CMSA Technical Report: 96-03

CHARACTERIZING RISKS IN EMERGING SOIL REMEDIATION TECHNOLOGIES

WORKING PAPER SERIES

CENTER FOR MODELING, SIMULATION, AND ANALYSIS
DEPARTMENT OF OPERATIONAL SCIENCES

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

SCHOOL OF ENGINEERING
Wright-Patterson Air Force Base, Ohio

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
March 1996

3. REPORT TYPE AND DATES COVERED
Technical Report

4. TITLE AND SUBTITLE
Characterizing Risks in Emerging Soil Remediation Technologies

5. FUNDING NUMBERS

6. AUTHOR(S)
Capt Thomas J. Timmerman, USAF; Dr. Richard F. Deckro;
LTC Jack M. Kloeber Jr., USA; Lt Col Jack A. Jackson, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
2750 P Street
WPAFB, OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
CMSATR/96-03

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of Energy/EM-50

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The Department of Energy is focusing a long-term development effort on producing cheaper, safer, and faster state-of-the-art soil remediation technologies. To assist with the management of these innovative technology development projects, ways of quantifiably measuring technical risk were investigated through a detailed literature review. "Technical risk" was defined in this study as the combination of the consequences of undesired events and their likelihood. Careful design of the inputs into a technology selection decision support system accounted for the uncertainty in forecasting final characteristics of remediation technologies still in the early phases of R&D. Experts made subjective probability estimates of these cost, schedule, and performance factors. Examination of several measures of final cost and schedule risk focused on communicating the risks inherent in different technological alternatives to the technology manager for operational, not theoretical, use. These risk measures included subjective measures, using utility theory, and objective measures, using variation about an expected value. A new measure was developed, the expected unfavorable deviation, which is similar but superior to the semi-variance as a measure of downside risk. These simple risk measures can be used whenever uncertainty is expressed through probability distributions of cost, schedule, and performance characteristics.

14. SUBJECT TERMS
Decision Analysis; Life-Cycle Cost; Risk; Technology Evaluation; Risk Analysis

15. NUMBER OF PAGES
244

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

CMSATR/96-03

CENTER FOR MODELING, SIMULATION, AND ANALYSIS
TECHNICAL REPORT 96-03

Presented to the Department of Energy (DOE EM-50)

by the Center for Modeling, Simulation, and Analysis

Department of Operational Sciences

Air Force Institute of Technology

Air University, Air Education and Training Command, USAF

Authors: Capt Thomas J. Timmerman, USAF; Dr. Richard F. Deckro;
LTC Jack M. Kloeber Jr., USA; Lt Col Jack A. Jackson, USAF

CHARACTERIZING RISKS IN EMERGING SOIL REMEDIATION TECHNOLOGIES

MARCH, 1996

TABLE OF CONTENTS

LIST OF FIGURES

Figure | Page

Figure	Page

Figure Page

Figure	Page

LIST OF TABLES

Table Page

Table Page

Abstract

The Department of Energy is focusing a long-term development effort on producing cheaper, safer, and faster state-of-the-art soil remediation technologies. To assist with the management of these innovative technology development projects, ways of quantifiably measuring technical risk were investigated through a detailed literature review. "Technical risk" was defined in this study as the combination of the consequences of undesired events and their likelihood. Careful design of the inputs into a technology selection decision support system accounted for the uncertainty in forecasting final characteristics of remediation technologies still in the early phases of R&D. Experts made subjective probability estimates of these cost, schedule, and performance factors. Examination of several measures of final cost and schedule risk focused on communicating the risks inherent in different technological alternatives to the technology manager for operational, not theoretical, use. These risk measures included subjective measures, using utility theory, and objective measures, using variation about an expected value. A new measure was developed, the expected unfavorable deviation, which is similar but superior to the semi-variance as a measure of downside risk. These simple risk measures can be used whenever uncertainty is expressed through probability distributions of cost, schedule, and performance characteristics.

CHARACTERIZING RISKS IN EMERGING SOIL REMEDIATION TECHNOLOGIES

I. *Introduction*

1.1 *General Issue*

Technology planning is an essential function for any government or private organization involved with investigating and procuring new matériel. Motivated by competition in the marketplace or concerns of national security, new technology is sought as a response to changing requirements. Advances in technology are also pursued to meet needs that currently go unsatisfied. Successful organizations must balance the opportunities offered by new technologies against the costs of researching and developing them. This is particularly true when one considers how a firm may invest considerable time and effort in research and development only to find the results insufficient to justify the expense. New technologies can be directly investigated by the interested organization or found outside in the marketplace, but any organization that wishes to survive and thrive must constantly assess emerging new technologies for eventual future application and/or impact, trading off today's resources for future capabilities. Unfortunately, when dealing with the state-of-the-art, these future capabilities are by no means certain. The development of new technology is inherently risky.

There is always some risk involved with strategic and tactical R&D decisions — risk that the technology will not be ready at the time it is required, risk that it will not perform as predicted, risk that the development costs will be higher than anticipated, and so forth. One must gain some insight into both the likelihood of these difficulties occurring and their consequences to intelligently invest an organization's resources in settings of less than certainty.

The nature of emerging technology hinders such assessment. Predicting the success of an R&D effort or the eventual performance of some new manufacturing process or weapon system is a formidable task under the best of conditions. While in some cases one can extrapolate future capabilities from past development efforts (e.g. Moore's Law: the number of transistors and therefore the computing power of microprocessors doubling every eighteen months [Bronson, 1996:192]), for products involving innovative technological approaches which are fundamental shifts in capabilities there are often no historical data to draw upon. Generally in such cases one must resort to the enlightened speculations of those with special in-depth knowledge and expertise in the specific subject to predict the eventual results of research and development efforts [Millett, 1991:43].

One such area of research and development is in the remediation of buried hazardous, often radioactive, waste. Although positive steps have been taken during the past thirty years to remedy the nation's environmental problems, many environmental and economic challenges remain. To answer these challenges, the U. S. Department of Energy

(DOE) has been implementing an aggressive national program of applied research that encourages the development of technologies to meet environmental restoration and waste management needs, focusing on the DOE's most pressing major environmental management problems. The keystone of the DOE's approach is to develop remediation technologies that are better, faster, safer, and more cost effective than those currently available [DOE, 1995a:vii-viii]. These innovative technological approaches lie at or near the frontier of the state-of-the-art. Due to the innovative nature of many of these projects, the DOE lacks historical experience upon which to base forecasts. As these technologies progress toward eventual employment, the DOE will be driven by limited budgets to fully fund only the most promising approaches. Obviously technology forecasting is of crucial importance to these decisions, despite the difficulties involved.

The stakes involved in waste remediation and environmental protection are high. The extent of the waste remediation problem facing the United States is enormous. There are 3.1 million cubic meters of buried waste on DOE installations alone, with an associated 40 million gallons of contaminated ground water [Mohuidden, 1995b]. The US Environmental Protection Agency has listed over 1300 Superfund sites across the country that must be cleaned up [Luftig, 1995]. The remediation of these waste sites will require the support of a long-term research and development program to identify lower cost alternative approaches to currently established techniques. To date, many remediation methods have been unsuccessful, difficult to implement, or exceedingly costly [Rumer, 1995]. Historically, these methods have included waste containment in barrels, concrete

blocks, and geologic repositories [Jackson, 1995:1]. The total life cycle costs of these clean-up efforts could potentially exhaust the nation's ability to pay for them, over the seventy to a hundred year time span the national program is expected to last [Mohuidden, 1995a]. Over **$750 billion** will be spent on remediation in the U.S. in the next thirty years alone [Gilliam, 1995]. Both the costs involved and the long-term nature of the national remediation program demand careful technology planning to minimize the financial and environmental burden of future generations of Americans.

1.2 *Background*

1.2.1 *Risks Involved in Technology.* The Department of Energy, like many other organizations, must develop new capabilities to meet current and future requirements. But to truly succeed, the DOE has to "win the gamble" by investing in technologies that payoff in the needed capabilities. Risk is implicit in the decisions made by DOE management, because the eventual outcome of an R&D effort is uncertain until the project is completed and deployed in the field.

To a program manager, risks are all in relation to delivering a specified product or level of performance at a specified time for a specified cost. A wide variety of problems and events can prevent the meeting of these cost, schedule, and performance objectives [DSMC, 1989:3-3]. The anticipation of failing to meet these goals forms the risk in the program.

"Risk" is a difficult term to use precisely. Common meanings of the word include the chance of injury, damage, or loss and a hazard or dangerous chance. By this usage,

anything with a possible undesired or unfavorable outcome has risk. The ambiguity between risk as the likelihood of the undesired event and the event itself keeps precise definition difficult. The "chance" of an harmful event reflects the uncertain future.

In practice, the difference between the terms "risk" and "uncertainty" is often obscured. Although managers in both financial and technical fields often confuse these two concepts [Bhat, 1991:262], in program management "risk" is often taken to mean the likelihood of an unfavorable event happening and the significance of the event's consequences. The term "uncertainty" describes how the ultimate outcomes of the project are unknown, and so deals with the likelihood of events and not events themselves. To truly understand whether a potential event is risky, one must have an understanding of the impact of its occurrence (or non-occurrence) [DSMC, 1989:3-1].

While there are other sources of program risk, including management difficulties, funding delays, and other environmental effects, a great deal of risk can be associated with the technology being developed itself. The attempt to provide a new or greater level of performance than previously demonstrated, or a similar level of performance subject to some new constraints of budget, packaging, or time, carries with it the possibility of failure with the consequence of wasted time and money. This risk is generally referred to as "technical" or "technological risk," and is of critical importance to projects trying to improve on the state-of-the-art [DSMC, 1989:3-3].

Concept of Risk

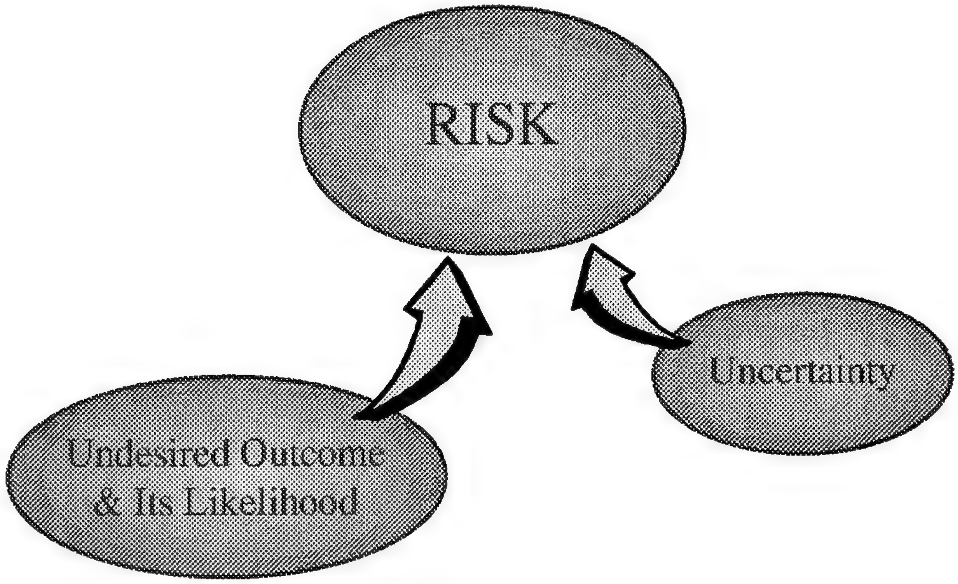


Figure 1.1

For the moment, then, let our concept of technical risk be the combination of unfavorable events springing solely from the technology that impact cost, schedule, and performance objectives with the likelihood of their occurrence, together with the uncertainty involved with not knowing what will actually occur. We will refine this definition after examining several different ways of quantifying risk in Chapter II. Estimating technological risk, however, is problematic. Figure 1.2 graphically depicts the categories of knowledge with which the manager must deal. Known data are readily available to the planner. Knowable data are those that can be collected by investigation, testing, program reviews, or other established methods. Unknowable data cannot be ascertained at the current point in time, most often because they depend on future results. The degree of uncertainty increases as one goes from the known to the unknowable. As

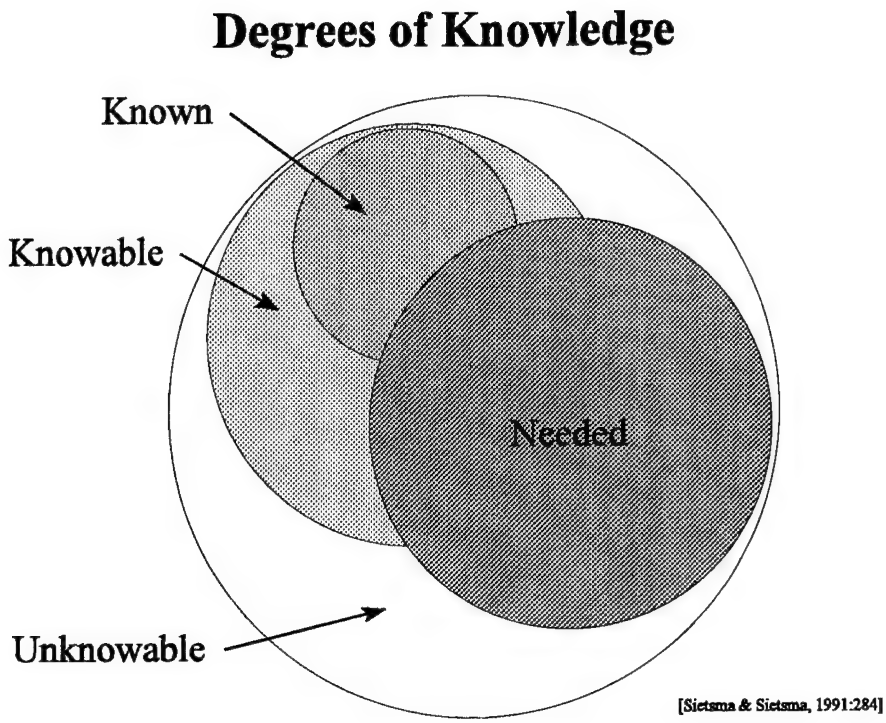


Figure 1.2

the figure suggests, the necessary information to understand the risks involved comes from all three categories. While possible events can be anticipated, the actual probabilities of their occurrences lie in the unknowable category and therefore must be estimated and/or approximated.

Unfortunately, one common way to deal with uncertainty in analyzing program and project management is to ignore it, conducting business as though current projections are 100% accurate. The underlying assumptions are that the project is deterministic and all factors are knowable, and that planning could be made practically watertight if only time and resources allowed development of sufficient detail in the plan [Sietsma and Sietsma, 1991:284]. This is a poor way to serve technology decision makers.

"Ignoring the inherent variation or uncertainty only masks its effects and give an unwarranted veil of pseudo-accuracy to the analysis. Furthermore, if the total uncertainty is significant, not recognizing it will often totally distort the results of the analysis in an unknown way, making any decision based on the analysis highly suspect" [Choobineh and Berhens, 1992:907].

Inclusion of the risks involved is therefore an important part of helping program managers make technology investment choices.

1.2.2 *The Office of Technology Development.* The sponsor of this study, the Office of Technology Development (EM-50), has the mission of researching new and innovative technologies to meet the DOE's environmental remediation needs. EM-50 works with other programs within DOE, other federal agencies, national labs, universities, and the commercial sector to maximize research efforts and ensure safe and efficient clean-up. Its goals are to develop technologies that make remediation safer, more cost-effective, and compliant with existing regulatory requirements. In many cases, development of new technologies presents the best hope for ensuring a substantive reduction in risk to the public, the workers, and the environment [DOE, 1995c:4].

The primary customers of EM-50 are two other major parts of the Environmental Management division of the DOE. The Office of Waste Management (EM-30) is responsible for treating, storing, and disposing of waste, and managing spent nuclear fuel generated during weapons processing and manufacturing, research activities, and site remediation activities. Currently, DOE facilities house more than one million cubic meters of radioactive waste. EM-30 is also responsible for coordinating waste minimization and pollution prevention efforts for the entire DOE. The other primary customer of EM-50 is the Office of Environmental Restoration (EM-40). Their mission is to protect human

health and the environment by remediating contaminated soil, groundwater, surface water, structures, and other materials at EM sites. Other EM-40 responsibilities include necessary landlord, oversight, surveillance and maintenance, and technical assistance to support remediation work [DOE, 1995c:2-4].

1.2.3 *Department of Energy waste remediation responsibilities*. The Department of Energy is responsible for cleaning up approximately 3.1 million cubic meters of buried waste at various landfills on government property throughout the U.S. This waste is predominantly located at six DOE installations: Hanford, Savannah River, the Idaho National Engineering Laboratory (INEL) at Idaho Falls, Los Alamos National Laboratory, Oak Ridge (X-10), and Rocky Flats. About half of this waste was buried before 1970, predating the more strict environmental regulations of the past three decades. Previous disposal regulations permitted the commingling of various types of waste; therefore, much of the buried waste throughout DOE sites is presently believed to be contaminated with both hazardous and radioactive materials (so-called mixed waste), a situation which greatly complicates remediation efforts (see Table 1.1 for types of waste [DoD:1994, 2-1]).

Typical buried waste includes construction and demolition equipment (such as lumber and concrete blocks), laboratory equipment, processing equipment (such as valves, ion exchange resins, and particulate air filters), maintenance equipment (such as hand tools, cranes, and machine oils), and decontamination materials. Typical disposal containers included steel drums of various sizes, cardboard cartons, and wooden boxes.

Larger individual items were disposed of separately as loose trash. Degradation of the

Types of Waste
Volatile organic compounds (VOCs)
Semivolatile organic compounds (SVOCs)
Fuels
Inorganics (not including radioactives)
Explosives
Low-level radioactive waste (LLW)
Low-level mixed (radioactive and hazardous) waste
High-level radioactive waste

Table 1.1

waste containers is believed to have resulted in the contamination of the surrounding soil as well [DOE, 1995b:6]. Since more than twenty five years has passed since much of the waste was buried, in some cases no documentation of exactly what was buried has survived [Mohuidden, 1995a].

The resulting uncertainty of exactly what waste types and items exist in a given landfill complicates the remediation process. Even a technology that has proven itself reliable and effective at other sites may "fail" when an unanticipated waste stream is found that the technology is incapable of effectively handling. Thus the first step in any remediation process is a careful assessment of what waste lies beneath the surface of the landfill (see Figure 1.3). This characterization and assessment is also a potential source of

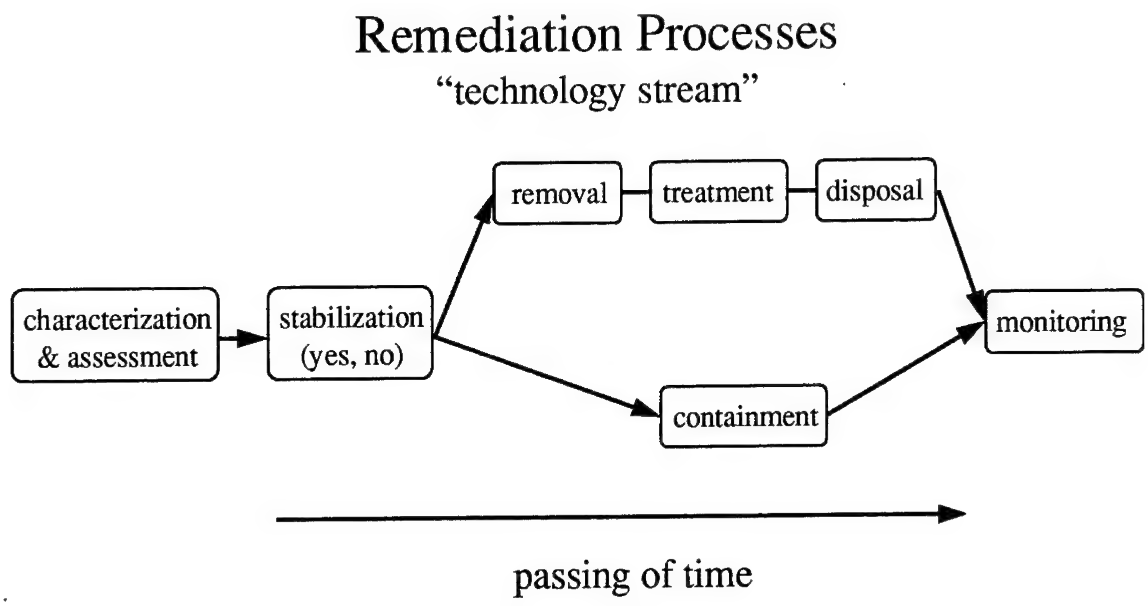


Figure 1.3

uncertainty, as the characterization may not be accurate or precise.

When the characterization is sufficiently complete, the major decisions of how to remediate the landfill must be made. In general, there are two approaches: 1) removal of the waste from the ground, followed by some treatment to make the waste manageable, and then storage of the treated waste (either on or off site); or 2) containment of the waste on site behind some sort of "barrier" which prevents further leaking of the waste into the surrounding environment. Temporary stabilization of the waste stream may also be used to prevent waste from reaching the environment until some more permanent solution is implemented. The use of one particular approach is not exclusive — different characterization, treatment, and/or containment technologies may be combined during one clean-up to cover different waste types in a "treatment train." The final stage of any

remediation is the placement of monitoring stations around the landfill and/or the waste storage location to watch for waste that might have been missed or degradation of the containment system [Mohuidden, 1995a].

1.3 *Report Scope and Organization*

This study investigates means to incorporate quantitative and qualitative risk measures in examining emerging technology. This research had two principle objectives. The first was to develop part of a decision support system to aid the DOE in selecting landfill remediation technologies for further funding, based on life-cycle cost modeling and risk criteria. The model is being developed under contract to the DOE Landfill Stabilization Focus Area, as a cooperative effort of the Air Force Institute of Technology's Department of Operational Sciences (AFIT/ENS) and a DOE contractor, MSE Technology Applications Inc. This study concentrated on the technology risk characterization framework for this decision aid, combining ideas from risk assessment and technological forecasting literature. See Chapter III, section 3.1 for a detailed description of the decision support system. An ancillary goal of this study was to conduct a more general investigation of assessing the risks of emerging technologies.

1.3.1 *Scope.* This research focused on soil remediation technologies, with particular attention to the technologies demonstrated as part of the DOE Landfill Stabilization Focus Area projects. The specific risk factors that are examined through the technical risk assessment framework are listed in Table 1.1 below. These risk factors

Risks Assessed in Technical Risk Characterization Framework

risk in...	method	used by
development schedule	distribution of dates when technology completes R&D	LCC Module
development costs	uniform cost per year of R&D	LCC Module
implementation performance	probability that technology will work successfully in the field	Decision Analysis Module
compliance with regulatory requirements	question user if the technology meets the regulation requirements governing the landfill in question	Technology Database (screening criteria)

Table 1.2

were selected by the project team in October 1995 to establish the information/communication requirements between the different modules of the overall model (see Figure 3.1).

This research concentrated on the process of estimating these risk factors. Information about the technologies assessed for demonstrating the overall model was provided by MSE. Since actual performance data for these emerging technologies was not available, reliance on expert judgements about the technologies' future capabilities was required.

Only a cursory treatment of the research and development costs of emerging technology was conducted in this study, as cost analysis is the research focus of the LCC modeling effort. Simplifying assumptions about the distributions of cost between different phases of the R&D process were made. To provide a detailed treatment of R&D cost estimating for each specific technology is outside the scope of this research. Such a study

would require a detailed engineering analysis of *each* individual system and its individual characteristics. A more general model, able to review a wide variety of remediation technologies, was the objective of this study.

1.3.2 *Report Organization.* The results of the literature review are discussed in Chapter II, while the methods used to estimate the technical risk factors are described in Chapter III. Also included in Chapter III are additional discussions of measures of risk that can be used to distinguish between recommended technology portfolios. The results of exercising these concepts on a set of demonstration technologies selected by MSE are discussed in Chapter IV, while conclusions and recommendations for further work lie in Chapter V. Preliminary computational results from the decision support system using notional technology data are included in appendices.

II. *Literature Review*

Since this thesis supports the development of a life-cycle cost and technology selection decision model to aid technology managers in making their technology investment decisions, this chapter will be organized by practical issues. Different ways to define and quantify risk will be discussed first, including ideas drawn from both risk assessment and technological forecasting literature. The special nature of innovative and novel technology complicates this definition, since there are greater uncertainties involved with assessing the technologies' characteristics. A discussion of risk analysis and technology forecasting and their use in program management follows. The nature of emerging technologies requires the use of subjective expert judgement, and therefore most of the remainder of this chapter is devoted to ways of soliciting and using expert opinion for assessing risk. Finally, some comments about public perceptions of risk will round out the literature review for this work.

2.1 *Concepts of Risk From the Literature*

While the Department of Energy has defined "risk" and "risk assessment" in its documents, it has taken "risk" to refer to only health and environmental issues. In a similar fashion as our general concept of risk formed in Chapter I, the DOE says risk is "the probability that something will cause injury, combined with the potential severity of that injury" [DOE, 1995c:67]. For the moment, let us distance ourselves from a specific definition and consider several different concepts of "risk." The definition we use for

"risk" sets the form we use to quantify and measure it, and so this definition should be selected carefully. One used in this study may not be appropriate for some other later risk analysis, and so this issue should be re-examined at the beginning of any study. The selection of a "measure of effectiveness" must be done with careful thought [Attaway, 1968:55].

2.1.1 *Qualitative Assessment of Risk.* Having said that our objective is to qualitatively assess risk, we should mention that qualitative rankings are often used. One way that is often used to characterize the risks of different alternatives is to use subjective judgement to give each alternative a "risk score," using some kind of qualitative numerical scale. This simple way of assessing risk bypasses the difficulties of objectively measuring it and can quickly produce results from a panel of experts or the decision maker.

Ryan states in an article dealing with assessing risks of new technologies that "some form of sophisticated numerical risk rating" is unnecessary for associating risk with technologies. Once technologies have been identified as part of a project, all that is required is "simply classifying [their] risk as low, medium, or high." Low risk technologies are not expected to present problems if traditional practices are followed. Medium risk technologies require special measures during development to "ensure that [development] proceeds properly," while high risk technologies may fail even with "special measures" [1990:69-70].

A similar approach was used in a recent study of different treatment technologies that use thermal mechanisms in their process. There, using topics established in the

Federal Facilities Compliance Act of 1992, experts qualitatively assessed scores of the different alternative technologies using high, medium, and low levels. Some of these topics included total LCC, environmental and heath risks, and risks of regulatory compliance [Feizollahi and Quapp, 1992:5-1, 5-41-3].

The difficulty in this approach is that "risk" is often not specifically defined. Making trade-offs between risk and other decision making criteria is difficult, since objective relationships between the criteria are not known. What is "high" for one person may be "medium" to another. While these and other problems exist with subjective and qualitative assessment, this sort of categorization of technologies is quick and may be all a decision maker requires. In our problem, however, more quantitative measures are desired.

2.1.2 *Ways of Dealing With Uncertainty.* If we are going to quantify risk, we must start with the concept of uncertainty. Uncertainty about the actual outcome of a future event with the potential for undesirable consequences is part of our concept of risk. Uncertainty reflects a lack of knowledge about the true state of events. One may lack knowledge about both the chance and the consequence of an uncertain event. If there was no uncertainty, there would be no risk. The outcome would be known and determined.

It is useful to distinguish between not knowing what the potential outcomes of a "risky" event are and not knowing which of a set of known outcomes will actually come to pass. Helton labels these states of knowledge as "subjective uncertainty" and "stochastic uncertainty," respectively. Analysts traditionally express subjective uncertainty

through establishing a set of possible outcomes and using probability distributions to characterize where the true outcome lies in that set. Examples from project management would include predicting a product's final delivery date or total development costs. Stochastic uncertainty, on the other hand, is addressed by examining the totality of possible outcomes and their likelihood of occurrence. More information is known under stochastic uncertainty than with subjective uncertainty. Helton also describes "completeness uncertainty," where the question is raised of including all of the possibilities inside the boundaries of the modeled set of potential outcomes [1994:483-6]. Application of the completeness uncertainty concept is difficult, since we cannot know what we do not know, but can be used with subjective feelings of confidence (see section 2.1.3 below). Emerging technology management deals more with subjective than stochastic uncertainty, and so that is what will be meant by "uncertainty" in the rest of this text unless specified otherwise.

2.1.2.1 *Subjective Probability*. The basis of the above definitions of uncertainty is the concept of probability. While many introductory statistics textbooks introduce "probability" as a relative frequency of a certain outcome occurring over a long term period [Mendenhall, et. al., 1990:17-8], this definition is of little use in the case of innovative technological R&D. Many of the events of interest happen only once: for example, the completion of a specific research program, the success or failure of a given field test, or the signing of the final government payment receipt for a particular item. Thinking in terms of long-run frequencies or averages makes little sense for one-of-a-kind

events, and so a different view of probability will be used. Looking at Figure 2.1 below, we can see the contrast between traditional objective probability and subjective probability — how more certainty is required for objective descriptions of probability. For our purposes, subjective probabilities will represent a degree of belief that an event will occur. There are no correct answers when it comes to subjective judgement — an event judged to be highly improbable may still happen without nullifying the original judgement. Without a sufficient number of identical trials, the validity of a subjective probability estimate cannot be verified [Clemen, 1991:208-10].

These subjective probability estimates are traditionally used to represent subjective uncertainty in simulation and decision analysis. The set of possible events and assigned probabilities can be used to find expected values of the parameter in question. The

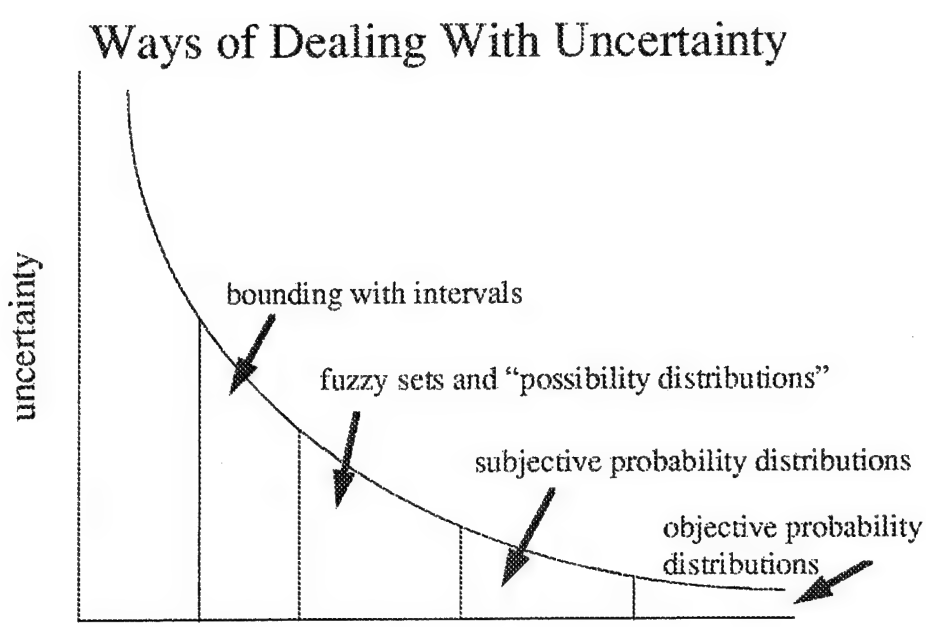


Figure 2.1

expected values of unknown variables are often used instead of known coefficients for deterministic math programming approaches to dealing with risk [Weber, et. al., 1990].

The point below or above which the actual value of the parameter will fall can be found from the cumulative distribution for some set probability. This is useful in reliability studies, where comparing the times where, say, 1% of a set of sub-systems will fail is a key criterion for choosing which type of sub-system to buy. Establishing these probability distributions can be difficult. Attempts should be made to obtain the highest quality estimates practical, but the fundamental difficulty of predicting the unknowable remains.

2.1.2.2 *Intervals and Bounds*. As Figure 2.1 shows, using subjective probability to describe unknown parameters does require some certainty, either in prior knowledge of the parameter in question or assumptions in order to settle on the type of probability distribution to use for the estimation. If assumptions cannot be justified or prior information does not exist in sufficient quantities, other methods may be necessary.

One approach that requires the least known or assumed information is to estimate the absolute limits of an interval which contains the parameter in question. For example, managers may try to estimate the time when a manufactured product will be delivered to a customer. They can bound the actual delivery date with the earliest and latest possible dates and form an interval.

These bounds can either stand on their own as a statement of what is possible and impossible for the estimated parameter in question, or be used in a model of some process to generate further intervals for other important variables. If one had interval estimates for

the n inputs of such a model, one would need to consider all possible 2^n combinations of these inputs to find the bounds on the output, an approach called the vertex method (named for the vertices of the n-dimensional feasible space for the model output) [Choobineh and Behrens, 1992:909-10]. The interval of possible values of the output would then be known, subject to the believability or the original input interval estimates and the model.

The usefulness of bounds is questionable, however. While interval analysis is relatively simple to use and requires the minimum level of information, the instantaneous transition at the bounds from possible to impossible can be a poor or counter-intuitive assumption [Choobineh and Behrens, 1992:917]. Another difficulty with intervals is assigning meaning to the bounds of results from interval arithmetic on other intervals. Say one was trying to find the bounds on the possible remediation costs for a landfill, using stabilization and a retrieval-treatment-disposal strategy and a known volume of low-level waste. The lower bound for the total cost would be the sum of all the lowest process costs, while the highest bound would the sum of the highest. Even knowing nothing about the way the costs are distributed for each process, one can see that it is very unlikely for the total costs to be at one of the bounds. If one takes a set of intervals as the limits of uniform or unimodal probability distributions, bounds on the sum or product of the set resulting from even mildly correlated input variables may represent likelihoods so low as to be practically worthless [Auclair, 1996]. However, knowing the upper and lower bounds (i.e. the best and worst cases) of a uncertain outcome can be valuable.

2.1.2.3 *Fuzzy Sets and Possibility Distributions.* An approach requiring an intermediate amount of certainty, falling in between subjective probabilities and intervals, is the use of fuzzy set theory. It is an extension of interval analysis to include multiple intervals with different levels of completeness uncertainty.

Instead of just one interval of possible values for the unknown parameter, successively smaller multiple intervals are established with the understanding that the value of the parameter is contained within the intervals with successively lower subjective probability. Possibility distributions (as opposed to probability distributions) act as the "membership function" of the parameter. The membership function of a level of the parameter indicates the degree of "belongingness" of that level in the set of possible values, and are often subjectively assessed through simple linguistic descriptions of sureness and certainty. Membership functions are expressed as being between 0 and 1. Using a threshold value, α, one can generate crisp ordinary intervals from the set of possible values by including only those levels that have a membership function of greater than or equal to α. This α is called "the level of presumption" and the resulting interval is called an "α-cut." Interval arithmetic can then be used to find output intervals for a given α [Choobineh and Behrens, 1992:911-2].

The definition of α requires that the possibility distribution be unimodal. If the membership function of the parameter value i is μ_i, where $\mu_i \in [0,1]$, the α-cut of the fuzzy set I is I_α, which contains all the possible values in I such that $\mu_i \geq \alpha$. A possibility distribution can then be constructed by a series of k nested intervals such that $I_{\alpha 1} \subseteq I_{\alpha 2} \subseteq I_{\alpha 3} \subseteq \ldots \subseteq I_{\alpha k} \subseteq I$, where $\alpha 1 > \alpha 2 > \alpha 3 > \ldots > \alpha k$. These possibility distributions can be somewhat triangular in shape such as in Figure 2.2, although they are not restricted to such shapes. The possibility distribution can be used in ranking different intervals of the

parameter with regard to a decision maker's value of the level of certainty that the

Possiblity Function of Total Cost, i

i

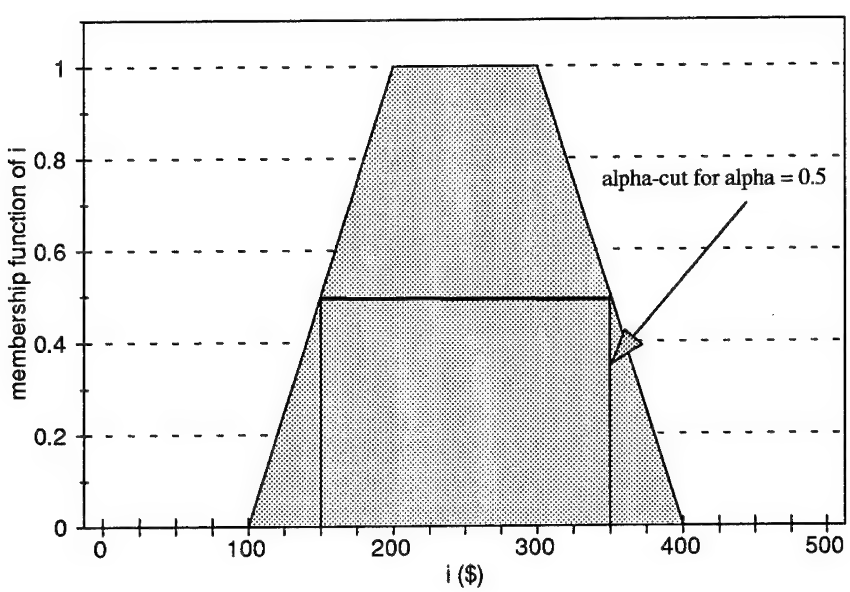


Figure 2.2

nterval contains the desired parameter [Choobineh and Behrens, 1992:911-15].

The creation of possibility distributions require less information about the parameter in question compared to subjective probability estimates [Choobineh and Behrens, 1992:908]. The "level of presumption," α, represents the likelihood that the estimated parameter will be contained in the interval, in a way that is very similar to confidence intervals developed in standard statistics. The level of presumption performs the same function as the confidence coefficient, the probability that the interval holds the parameter of interest [Mendenhall, et. al., 1990:353]. Possibility distributions are subjectively assessed confidence intervals, where expert opinion is used instead of statistics to define the bounds of the interval.

2.1.3 *Risk as a Probability and Associated Consequence.* The traditional approach in project management and risk assessment in defining "risk" and "uncertainty" is to use "risk" in situations that Helton would label stochastically uncertain, where the potential outcomes are known and only the probabilities of their occurrence must be investigated, and "uncertainty" where Helton would use "subjective uncertainty" [Bhat, 1991:262; Levy and Sarnat, 1990:190]. This difference is sometimes used to establish a border between what can and cannot be modeled, since "uncertainty" prevents clear knowledge of possible events. This is not a very useful distinction for us, since we are dealing with subjectively uncertain issues with emerging technology. One can postulate certain outcomes and proceed from there, building a worthwhile model of "uncertain"

events while keeping one's assumptions in mind. For the purposes of this study, Helton's terms are much more useful.

Formal Department of Defense guidance in program management defines "risk" as the likelihood of an undesirable event occurring and the significance of the event's consequences. "Uncertainty" addresses only the likelihood. To truly understand whether a potential event is "risky," one must have an understanding of the impact of its occurrence or non-occurrence [DSMC, 1989:3-1]. This approach may be more practical than that of the traditional project management definitions above.

The separation of risk into probability and consequence has other advantages, as well, by allowing risk control efforts to be split between *prevention* and *mitigation*. Prevention efforts are any set of actions that reduce the probability of undesired events, while mitigation efforts reduce the level of unfavorableness of an event. Prevention actions are not necessarily exclusive from mitigation efforts. In a sense, when using risk as a decision criteria for our remediation technology investment problem, we are evaluating different prevention and mitigation alternatives [Sherali, et. al., 1994:200]. If we compare future technologies to what currently we use in terms of, say, cost, prevention and mitigation would be expressed in the shape and location of the technologies' cost distributions.

As already discussed in section 2.1.2.1, subjective probability distributions are traditionally used to describe situations of subjective uncertainty. If the events in question

include unfavorable outcomes, then all the information needed to satisfy the DoD definition of risk is at hand once these probabilities are known or assumed.

2.1.4 *Concepts of Risk From Financial Literature.* Financial methods to deal with risk and uncertainty are often applied to evaluations of new technology. Ways of dealing with risk factors for evaluating different economic options have been proposed and used. Ignoring the uncertainties entirely is sometimes done [Choobineh and Berhens, 1992:907], but is only sensible when all the possible options are low risk to start with.

2.1.4.1 *Net Present Value.* Cash flow based methods such as net present value (NPV) and internal rate-of-return (IRR) are traditional tools of financial analysis of capital investments. Estimating NPV of the costs of an alternative requires both estimates of the cash flows and their timing, as one can see from Equation 2.1. This shows how to calculate the NPV of a stream of cash flows $x_0, x_1, \ldots, x_n$ over n periods, using an interest rate of i [Clemen, 1991:24-5].

$$\begin{aligned} NPV &= \frac{x_0}{(1+i)^0} + \frac{x_1}{(1+i)^1} + \ldots + \frac{x_n}{(1+i)^n} \\ &= \sum_{j=1}^{n} \frac{x_i}{(1+i)^j} \end{aligned} \tag{2.1}$$

The interest rate i (also called the discount rate) is chosen to represent the return one gets from the next best investment opportunity. NPV, then, is used as a relative measure of return on investment by comparison to some more certain rate of return. The choice of i is often used to reflect the riskiness of investments, by deflating the potential benefits of

alternatives judged to be "risky" in comparison to other options [Levy and Sarnat, 1990:245].

The IRR is the interest rate required to generate a NPV of 0. This is taken to be a more absolute measure of an investment's return, since different alternatives can now be compared to see what sort of equivalent certain return would produce the same net profit VanHorne, 1971:55]. Equation 2.1 is solved for i, resulting in an n^{th} degree polynomial that could have up to n real roots [Cain, 1996]. The difficulty with IRR is discriminating between the set of real solutions to find the "right" one [Levary and Seitz, 1990:31]. There may only be one positive real root, but if there are multiple feasible roots there are no ways to judge which is "right." For this reason IRR is not always an appropriate measure of financial risk [Cain, 1996].

Arguments against using NPV and IRR measures of technology risk include comments that they undervalue new technologies, because of the discounting effects of the calculations. Future benefits (represented by some positive cash flow) are given little weight compared to near-term net profits. NPV also requires a static view of future industrial activity, represented by the single interest rate. Many benefits that cannot be quantified in terms of money are ignored [Mitchell, 1990:155; Ashford, et. al., 1988:637-8].

A "hurdle rate" is sometimes set as an arbitrary expected rate of return or performance below which candidate projects are disregarded. It is based on the principle that high returns should follow high risk. This rule ignores the variance of the risk factors

around the expected value, and naively expects that demanding high expected performance will always produce high actual performance. Another approach is to adjust estimates coming from analysis groups by some historical average correction, accommodating the risk of poor estimation by adjusting their figures by some percentage increase or decrease derived from what would be needed on average to correct their past estimates. This ignores the variance involved with the groups' estimates [Troxler and Schillings, 1993:30].

Sometimes NPV yields poor results because the discount rate is set too high, exaggerated by several over-estimation tendencies that bias NPVs against long-term rewards. Ashford, et. al., argue that the error lies in unrealistic interest rates, not in using NPV. "Risk free" rates from government bonds of similar value should be used, perhaps with some additional risk premium. They also argue that benefits that are traditionally difficult to quantify, such as re-use of flexible equipment in other projects, can be included with careful work, and that interactions between technologies assumed to be independent should be included as well. The baseline case, used to compare against future possible improvements, must be selected with care, since one can easily overstate this extrapolated status quo future without reflecting the effects of competitors' advancements [1988:637-9].

These financial standards are not easily used alone when the technology being developed does not generate revenue or directly mitigate expenses. However, they can be used at least to objectively compare alternatives based on cost.

2.1.4.2 *Risk as Variation From an Expected Value.* Uncertainties in both the cash flows and their timing must be accounted for in some fashion to use our basic concept of risk. Indeed, the financial community has generally not distinguished between "risk" and "uncertainty" [Levy and Sarnat, 1990:190]. Finance literature has understood "business risk" as being the relative dispersion of the net operating income of a firm [VanHorne, 1971:46]. For our problem of technology investment, this translates into concerns about the relative dispersion or variance of important decision criteria such as cost and time. Subjective probability distributions can be used to describe the random variables used to express these criteria when objective data does not exist [Levy and Sarnat, 1990:191]. Risk is then expressed by the variance of the estimated distribution of the decision variable around the expected value, and can be measured by the variance or standard deviation [VanHorne, 1971:46; Levary and Seitz, 1990:64].

A relative measure of risk is the coefficient of variation, defined as the ratio of the standard deviation to the mean. Larger coefficients of variation mean larger risk [VanHorne, 1971:46].

Another related measure of risk is the semi-variance. It is calculated the same way as variance, but only including that part of the distribution in one direction above or below the mean. This measures "down-side" risk, when variation in only one direction is considered "risky" [VanHorne, 1971:186; Levary and Seitz, 1990:79-80]. The semi-variance is recommended for use when the PDF of the attribute in question is not

symmetrical and therefore the variance may misrepresent the risk of alternatives [Levary and Seitz, 1990:80].

One can use these different risk measures to characterize alternatives by both "profitability" or "costliness" and "risk," using the expected value and some measure of variation, respectively. Alternatives are compared on the basis of means and variance. If a choice has a better (higher or lower, depending) mean and a lower variance, it is clearly the preferred choice [Levy and Sarnat, 1990:214]. Other cases, where say one alternative has a better mean but a larger variance, require trading off "risk" versus "value" in some way.

Another approach using subjective probabilities is to use the resulting cumulative distributions to find the probability that the final decision variable will be above or below some target value. The alternatives can then be distinguished by their different probabilities [Levary and Seitz, 1990:64].

2.1.5 *Risk as a Perceived Characteristic.* Since there are many uses of risk in health, safety, project management, and military literature, it is possible to lose sight of an important practical issue while attempting to estimate occurrences and likelihoods — that the risk involved with a possible alternative is often a subjective assessment made by a decision maker or stakeholder, with an association of negative value that does not result from careful rational thought [Wheeler, 1993]. However risk is defined, its impact on decisions is through the preferences of the decision maker, whether those preferences are formed by intuition or by painstaking risk assessment. Analysis can describe known or

hypothesized risks, but ultimately it is the decision maker's values and trade-offs that express risk.

2.1.5.1 *Utility Theory.* Decision analysis (DA) methods traditionally treat risk implicitly by incorporating the decision maker's preferences. DA attempts to prescribe the best decision from a set of alternatives while addressing the inherent uncertainty in the situation and potentially multiple competing objectives, by maximizing expected utility.* Utility expresses the subjective values of the decision maker for various levels of an attribute [Clemen, 1991:2-3; Keeney and Raiffa, 1976:6].

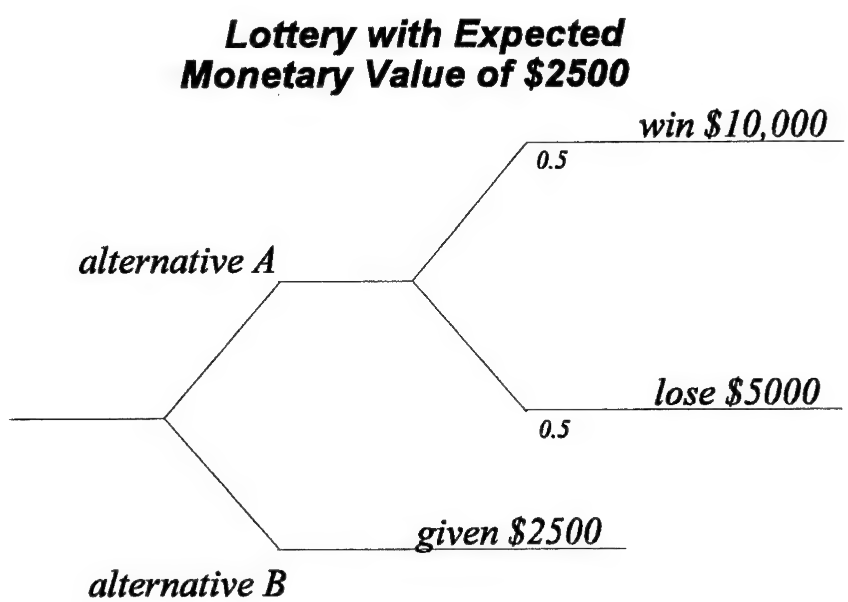


Figure 2.3

*In this thesis *utility function* always refers to a von Neumann-Morgenstern utility function used in decision analysis and multi-criteria decision making, not an economist's utility function [Keeney and Raiffa, 1976:150].

If a person is faced with a choice between two alternatives like the one shown in Figure 2.3, he or she may be indifferent between A and B since they have the same expected monetary return of $2500. Someone else may not feel the same, however, and take the certain $2500 rather than run the risk of losing $5000. A third person may forego the sure $2500 because the chance of winning $10000 is too appealing to resist. This difference in preferences, when the expected monetary value of the two alternatives are the same, is due to different feelings about the risk involved with alternative A [Keeney and Raiffa, 1976:149-50].

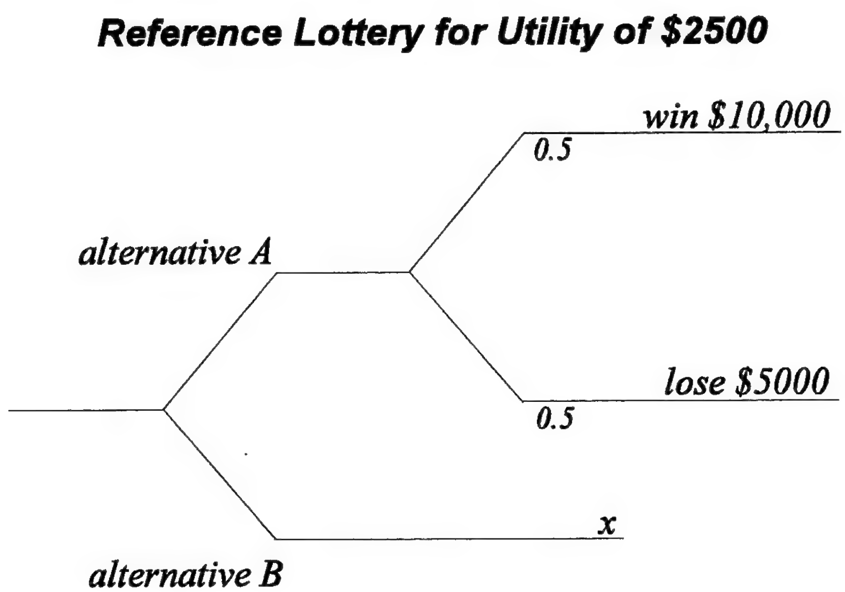


Figure 2.4

The way these feelings are captured for use in decision analysis is through utility functions, which mathematically express the subjective preferences of the decision maker. These utility functions are assessed using reference lotteries like that shown in Figure 2.4.

The same alternative A is used, which has an expected value of \$2500. A decision maker would be asked to examine this lottery and choose an x that would make him or her indifferent between alternatives A and B. If this x was \$2500, we would know that this person was neutral toward the risks of the gamble. If x was less than \$2500, we would know that he or she would prefer to avoid the risks, which we call "risk aversion." A "risk seeking" person would set x greater than the expected value [Clemen, 1991:367-7, 375].

The key point here is this: because the decision maker is indifferent between his or her x and the lottery in A, the utility of x must equal the expected utility of the gamble. If we know the utilities of winning \$10,000 and losing \$5000, we can average them to find the utility of x [Clemen, 1991:377].

Utility is measured between 1 and 0. We can set the utility of \$10,000 to be 1.0 since it represents the most money we could ever win, while the utility of -\$5000 can be 0 since it is the lower limit. Since the expected utility of alternative A is 0.5, we now know that the utility of x is 0.5 as well. We can now change alternative A to be a gamble between x and \$10,000 and find the new dollar amount that the decision maker is indifferent to, knowing that this will have a utility of 0.75. This can be repeated until the entire utility function is defined over the range [-\$5000, \$10,000].

This iterative procedure, using a general reference lottery like that of Figure 2.5, uses the concept of the *certainty equivalent* to piecewise assess a decision maker's utility function. In our previous example, x represents the guaranteed amount of money that has

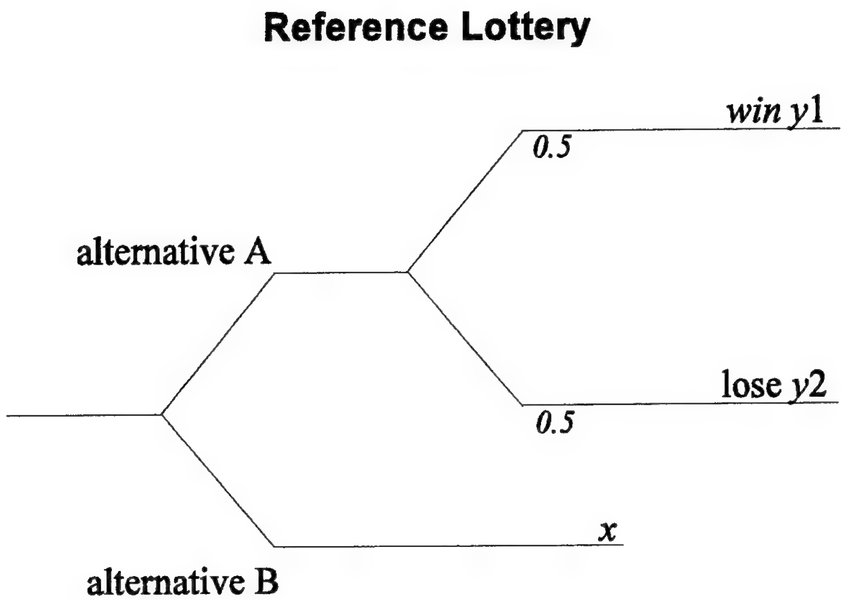


Figure 2.5

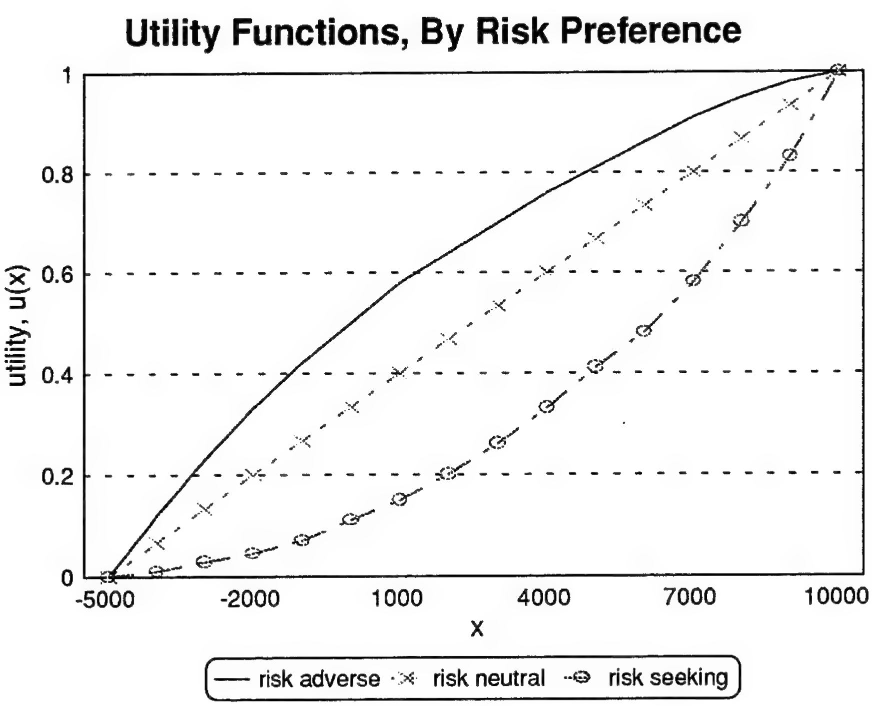


Figure 2.6

the equivalent value as the uncertain lottery in alternative A. This x is the certainty equivalent of the lottery in A, and will always be less than the expected monetary value for a risk averse person or more than it for a risk seeking one. The difference between the certainty equivalent and the actual expected monetary value of the lottery is called the *risk premium* [Clemen, 1990:371].

The *risk preference* is captured in traditional decision analysis by the shape of the utility function. Using reference lotteries like that in Figure 2.5 produces utility curves similar to those shown in Figure 2.6. For increasing utility functions, the concave utility function represents risk aversion, the linear function represents risk neutrality, and the convex function represents risk seeking preferences [Clemen, 1990:367-8].

2.1.5.2 *Risk as Marginal Utility.* A formalized version of the previous statement provides a measure of risk aversion through the following *local risk aversion function*, $r(x)$, defined on the utility function $u(x)$:

$$r(x) = -\frac{u''(x)}{u'(x)} \tag{2.2}$$

where $u'(x)$ is the first derivative of $u(x)$ with respect to x and $u''(x)$ is the second derivative. If $r(x)$ is positive for all x, $u(x)$ is concave and the decision maker is risk averse. If r is negative for all x, $u(x)$ is convex and the decision maker is risk seeking (notice that the utility function must be continuously twice differentiable for this risk aversion function to be defined). If two utility functions $u_1(x)$ and $u_2(x)$ are compared and

$r_1(x) > r_2(x)$ for all x, $u_1(x)$ indicates more risk aversion than $u_2(x)$ [Keeney and Raiffa, 1976:160-3].

Using this risk aversion function as a measure of the decision maker's feelings for risk, it is possible to define sets of utility functions based on their risk behavior. For example, decision makers tend to be more risk neutral when the decision involves monetary amounts that are small with regard to their total assets, say as the manager of large government projects or the executive of a large corporation. For these decisions, expected monetary value may be sufficient [Clemen, 1991:368]. Many types of risk aversion are possible, whether it is decreasing, constant, increasing, or even proportional to the amount of wealth at risk. The type of risk aversion can restrict the form of potential utility function to only certain ones, making risk aversion a powerful first step in assessing a decision maker's utility function [Keeney and Raiffa, 1976: 165-179].

It is important to remember that utility functions are only models of individuals' attitudes toward risk. They are defined for a specific set of objectives and criteria for the moment they were developed. It is dangerous to broadly interpret these revealed preferences. DA uses utility functions to add risk considerations to otherwise objective criteria as a way to model subjective decision making. However, a person's feelings toward risky alternatives can be complicated and may depend on what is at stake, the context of the decision, and the time horizon [Clemen, 1991:368]. Use of utility functions requires the assumed adherence to utility axioms which may or may not be violated by the decision maker.

2.1.5.3 *An Extension of Risk as Variation From the Expected Value.* The concept of risk as variation from the expected value taken from financial literature and the idea that risk is something perceived by the decision maker can be combined. This is the strategy that Jianmin Jia and James Dyer use to explicitly trade off the "risk" of an alternative against its "value." They develop a "standard measure of risk" by using the expected difference between the potential outcomes of a lottery and the mean of the outcomes. If x is a random variable representing the outcome of a lottery whose possible outcomes are members of the non-empty set $\{X\}$ and $\bar{x}$ is the expected value of x, then a new random variable x' can be defined as the difference between x and its mean $\bar{x}$. This x' is called the "risk variable" of the "value" x and represents the potential outcomes distributed around the mean $\bar{x}$. Note the expected value of x' is zero [1993:4-7].

Just as a utility function can be assessed representing the utility of x with standard decision analysis methods, a utility function for the risk variable x' can also be assessed for the decision maker which represents his or her feelings for risk explicitly. This utility function, $u_r(x')$, is the equivalent of $u_r(x - \bar{x})$ [Jia and Dyer, 1993:6].

Instead of assessing a new utility function, $u_r(x')$, Jia and Dyer use the original utility function $u(x)$ to express the value of the deviations from the mean. They define a "standard measure of risk" as the following:

$$R(x') = -E\,[u(x - \bar{x})] \qquad (2.2)$$

where $E[u(x - \bar{x})]$ is the expected utility of the mean of the difference between x and its mean when using the original utility function assessed on x [Jia and Dyer, 1993:5-6].

Increasing $R(x')$ means decreasing preference, assuming risk aversion. This risk measure is independent of the original mean of x and can be used as a measure of perceived risk. The potential alternatives $\{X\}$ can be ranked in accordance with $R(x')$, just as with any other expected utility, as an independent criteria that is used with others to form a decision analysis policy [Jia and Dyer, 1993:5-7].

The use of such a risk measure can be illustrated with a simple example. Suppose there were two possible outcomes of a gamble, a and b, with expected outcomes $\bar{a}$ and $\bar{b}$. If a has more variation about its expected value than b, $R(a) > R(b)$. Then b would be preferred over a if this risk measure was the only criterion for evaluating the choices. One can include non-risk criteria in evaluating the alternatives, however, and explicitly trade-off "value" against "risk" using multi-attribute utility theory, since Jia and Dyer's "standard measure of risk" is independent of any expected value or certain payoff of a or b [1993:7, 9].

2.1.6 *Summary and Refined Definition.* We can see that there are many ways to define and quantify risk in the literature. Financial methods concentrate on uncertainty and probability distributions, using variation about an expected value to objectively represent the risk of alternatives. Larger variation or range in the distribution of decision variables means more risk. Utility theory takes risk measurement in a different direction, assessing the subjective preferences of a decision maker for risk in deciding between different options. Typically, our decision makers will be risk averse, preferring less uncertainty to more. It is possible to look at alternatives by separating them into measures

of value (e.g. expected value, utility) and measures of risk (e.g. variance, Jia and Dyer's standard measure of risk), using objective or subjective measures.

Our concept of risk from Chapter I includes both uncertainty and the likelihood and severity of possible unfavorable events. Probabilistic methods are best used to quantify the subjective uncertainty involved with innovative technologies. Expression of each technology alternative through probability distributions of key decision variables will describe the probability of getting undesired cost, schedule, and performance outcomes in a way that satisfies our concept of risk.

Our definition of technical risk, then, will be the probability and associated consequences of achieving undesired outcomes in our key decision criteria of cost, schedule, and performance, expressed through subjective probability distributions. The risk embodied in these probability distributions can then be measured in different ways as desired.

2.2 *Risk and Program Management*

2.2.1 *Risk Management and Risk Assessment.* There has been a large number of articles, reports, and books published over the past decades that deal with various aspects of risk. Just as different definitions of "risk" are used, the practice of dealing with risk has been labeled and categorized in many different ways. This has been a source of continuing confusion in the literature.

The DOE uses its own terms to refer to the way health and environmental risks are examined in doing its day-to-day business. These definitions include 1) risk assessment:

technical assessment of the nature and magnitude of risk; 2) risk characterization: final phase of risk assessment process that involves integration of the data and analysis involved in hazard identification, source/release assessment, exposure assessment, and dose-response assessment to estimate the nature and likelihood of adverse effects; and 3) risk analysis: methods of risk assessment as well as methods to best use the resulting information [DOE, 1995c:67-8]. Since this study deals with cost, schedule, and performance risk, however, we need to look elsewhere for useful terms.

A clear distinction between risk assessment, risk analysis, and risk management is not widely accepted in the literature. The Defense Systems Management College in the report *Risk Management: Concepts and Guidance* defines "risk management" as the overall umbrella title for the processes that identify and manage risk. The report identifies two basic stages: planning and execution. Figure 2.7 shows the breakdown of their terminology [DSMC, 1989:4-1-2].

DSMC Risk Management Terminology

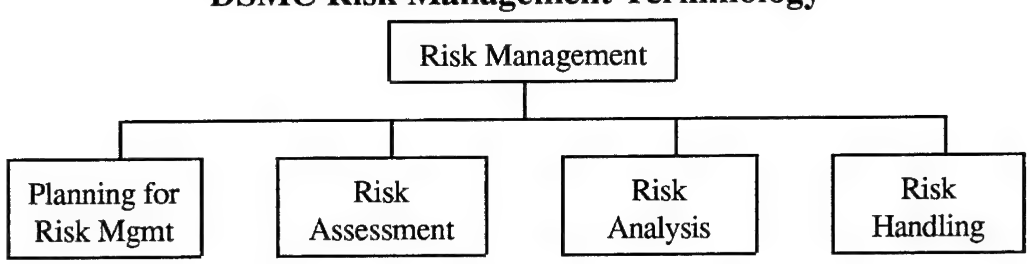


Figure 2.7

The purpose of risk management planning is "to force organized purposeful thought to the subject of eliminating, minimizing, or containing the effects of undesirable

occurrences" [DSMC, 1989:4-3]. This should be part of the overall planning begun before the program is initiated, including an integrated program schedule, and the resulting "risk management plan" should be updated as a matter of course during the program life span. The intended approach to identifying, assessing, analyzing, and handling the risks in the program should be laid out in this planning stage and kept current [DSMC, 1989:4-3-4].

The execution phase of this suggested risk management scheme then turns to identifying and describing the risks to the program through interviews of experts, the construction of analogies and baselines, and examination of the program plans. This is part of what DSMC calls "risk assessment," which leads to the comparison of program strategies with regard to the identified and roughly quantified risks. This process is not clearly separate from "risk analysis," which is an examination of the change in consequences to the overall program or sub-program caused by changes in those factors influencing the risks (i.e. sensitivity analysis). More sophisticated mathematical tools are used in this element of risk management, and the results are used in direct support of the program's decision makers. The transition from risk assessment to risk analysis is gradual over time, as a program matures [DSMC, 1989:4-5-10].

The last element, "risk handling," is the action taken to address the issues identified and evaluated in the risk assessment and analysis efforts. Avoidance of higher risk choices, attempts to prevent the occurrence and mitigate the effects of undesired events, and attempts to share the potential consequences across organizational and government-

contractor lines are performed. The acceptance of some level of risk has be made by the program decision makers in balancing the risks with their associated costs of prevention [DSMC, 1989:4-10-13].

This study fits the risk assessment definition well. By examining the programmatic and performance characteristics of candidate remediation technologies, the likelihood and associated consequences of budget and schedule problems of the national remediation program will be identified, within the limits of the gathered project data. The overall model development sponsored by the Landfill Focus Area is part of its risk management planning, providing a tool for risk assessment in the early parts of their program.

2.2.2 *Technological Forecasting*. The term "technological forecasting" is generally used to denote forecasting techniques focused primarily on predicting technological change over the long term. Technological techniques require imagination combined with individual talent, knowledge, foresight, and judgement to these changes. Use of these methods requires an understanding of the factors involved with each situation and the need to adapt the method to that situation [Makridakis, et. al., 1983:637].

The most important things about any forecasting effort is that it be credible and useful to a decision maker. If it lacks utility for the decision-making process, it is a failure. The methods used to process the best available information must be clearly described, methodologically sound, replicatable, and logically consistent. Assumptions and the confidence that can be placed in the forecast must be understood by the decision maker [Porter, et. al., 1991:52].

Millett and Honton broadly define technology forecasting as "the process and result of thinking about the future, whether expressed as numbers or in words, of capabilities and applications of machines, physical processes, and applied science" [1991:3]. Other definitions include the process of predicting the future characteristics and timing of technology [Meredith and Mantel, 1995:711]. According to Millett and Honton, technology forecasting should ideally provide a forecast of the future technological environment, suggest alternative technology strategies to managers, and evaluate these strategies to see which will produce the desired results [1991:ix].

These forecasts are guides for future action. As such, their accuracy is unknown when they are produced. The time horizon of the forecast is the best determinant of accuracy — the shorter the time horizon, generally the more accurate the forecast. Even inaccurate forecasts can be valuable, if the lessons drawn from them by decision makers are useful [Porter, et. al., 1991:54-5].

Care must be taken with technological forecasting, however. Meredith and Mantel emphasize that it is most appropriate when applied to future capabilities, not the characteristics of specific devices [1995:714]. Since we hope to assess the characteristics and timing of specific technologies, we should heed this caution and proceed carefully.

2.2.2.1 *Quantitative vs. Qualitative Forecasting*. A distinction should be drawn between traditional forecasting approaches and what is required for our problem. The structure of the traditional, general univariate quantitative forecasting problem is roughly where we have past values, up to some time t, of a random process X_0, ..., X_{t-2}, X_{t-}

$_1$, X_t, and wish to forecast the value X_{t+m} which the process will assume at the future time $t+m$. In constructing the forecast $\hat{x}_{t+m}$ we are answering the following questions: 1) What class of random processes are we considering? 2) What general class of functions of $\{X_s; s \leq t\}$ are we considering for $\hat{x}_{t+m}$? 3) Having chosen the general class of functions, what criterion of the accuracy of the forecast $\hat{x}_{t+m}$ should we use to determine the explicit form of $\hat{x}_{t+m}$ as a function of X_t, X_{t-1}, ...? Different answers to the second question lead to different functional forms and usually to different forecasts. Given satisfactory answers to the three questions and the true value of X_{t+m}, the "optimal" forecast is uniquely determined assuming the covariance structure of the X_0, ..., X_{t-2}, X_{t-1}, X_t, is known [Priestly, 1974:152].

It is important to note that the assumption that the future is a continuation of the past can be unjustified. Quantitative forecasts (based on the above definition) are conditional based on the past data and these assumptions being true. This can be a dangerous assumption to use without a meaningful theory of cause and effect [Millett and Honton, 1991:7-8]. Although the relationship between future variables is expected to be the same as in the past, in fact the validity of these assumptions is doubtful, as the future rarely follows directly from the past. If it did, simple trend extrapolations would be fairly accurate forecasts — but it is precisely because they are usually not that more sophisticated means of quantitative forecasting such as regression, econometric models, and systems dynamics were developed. These latter techniques recognize that the world is more complicated than simple forecasting models allow [Millett and Honton, 1991:40]

Millett and Honton's view is that these quantitative forecasts are a very important set of tools, but that they may be overemphasized and overrated, especially when one considers that their basic assumptions are about as valid or invalid as the expert judgement used for more qualitative forecasting. They are best used for forecasting near term events of up to two years [1991:41].

2.2.2.2 *Classification of Technological Forecasting Techniques.* Millett and Honton break up technology forecasting into three distinct categories: trend analysis, expert judgement, and multi-option analysis [1991:3]. Other classifications include Makridakis, et. al., who break the field up into subjective, exploratory, and normative approached [1983:639] and Porter, et. al., who use categories of monitoring, expert opinion, trend analysis, modeling, and simulation [1991:93-7].

Millett and Honton's trend analysis is the same as the quantitative forecasting described by Makridakis, et. al. [1983] and the trend extrapolation of Meredith and Mantel [1995:714-21], being the projection of past trends into the future as described above. One specific technique that they describe which is relevant to our remediation technology selection problem may be the use of historical analogies. Simply put, this is studying historical data from other similar technology development efforts to draw useful inferences for the project in question. This presumes that relevant data exist [Millett and Honton, 1991:25-6].

Expert judgement is the "assertion of a conclusion based on evidence or an expectation for the future, derived from information and logic by an individual who has

extraordinary familiarity with the subject at hand" [Millett and Honton, 1991:43]. This fits with our general use of the term expert opinion. Makridakis, et. al., describe subjective assessment methods in similar terms. They point out that, due to the subjective nature of these methods, the reliability of the results is often questionable. Consequently such results are often stated in terms of probability distributions and intervals, rather than single point estimates [1983:639].

These experts should possess three important characteristics: substantive knowledge in a relevant field or domain, the ability to cope when faced with uncertain extensions of that knowledge, and imagination [Porter, et. al., 1991:203]. Porter, et. al., believe that forecasts made by groups of experts are so much safer than those produced by individuals alone that they recommend not using expert judgement at all unless a group of experts from the relevant fields can be identified and recruited. Individuals acting alone can make wildly inaccurate estimates [1991:94]. While including other experts in the process may help exclude errors, they introduce other problems that have to do with group behavior.

Millett and Honton's discussion of this form of forecasting, which includes interviews, questionnaires, and group discussion methods, is heavily cited in the section on gathering expert opinion below. They point out that all methods of forecasting and analysis, to some degree or another, involve expert judgement, whether it is one person's or a group's, whether it is expressed in numbers or in words. However, expert opinion becomes particularly important in the analysis of highly uncertain and complex topics such

as ours. Many successful managers trust their intuition, which must be of some service or else they would not be successful! These same managers can be very skeptical of other people's expert judgement and demand justification of it based on logic and information before they will easily accept it. Millett and Honton judge that expert opinion alone is not a very satisfying forecasting method, but that it is an excellent method of gathering information for use with other methods [1991:43-44, 61].

Multi-option analyses is different than the other two categories that Millett and Honton use, in that these techniques examine alternatives in multiple possible futures instead of trying to nail down the one single future that is actually coming. This distinction is due to the way multi-option techniques accept the fact that we can never know what the future will be with sufficient certainty, and so they estimate likely alternative futures and plan towards at least one of them. These "multi-option" approaches are typically used by organizations that face repeated and significant changes in their operating environments. Millett and Honton describe scenarios, simulations, paths/relevance trees, and portfolio analysis as multi-option analysis techniques [1991:63]. Scenarios are also mentioned by Meredith and Mantel and Makridakis, et. al., and may be applicable through hypothesizing a worst case future, a best case, and a future where current trends continue. Organizational, economic, political, and social variables should be included as well as technological ones [Meredith and Mantel, 1995:724-5].

Many of these multi-option procedures are not generally accepted as "forecasting" techniques, at least not by quantitative forecasters. Whatever they may be called, Millett

and Honton state that these methods are certainly strategic planning and analysis approaches that are used with more than just technology, and do well with relating technologies with non-technical factors [1991:63-5].

2.2.3 *Cost and Schedule Estimates.* While this study is not intended to examine cost estimating in detail, risks involved in estimating the development and implementation costs of innovative technology are crucial issues for technology managers. Examples from DoD experience may be illuminating, as the procurement of new military hardware is similar in some respects to the development of innovative remediation technology. Most new weapons and other equipment contain new, untried technology [Biery, 1986:14] that are often not transferable to the commercial world.

The structure of the defense industry and the way military equipment is procured leave little encouragement to defense contractors to deliver goods on time and within budget. Indeed, the manufacturers have every incentive to make highly optimistic cost and schedule forecasts in order to win contracts. The sponsors are also motivated to accept optimistic forecasts to convince Congress and their supervisors that the program can fit into this year's budget. After the contract is awarded, there are few mechanisms available to control costs and schedules, so extra costs and time must often be accommodated since the only other choice would be to cancel the program and start all over [Biery, 1986:14].

The technology manager must understand that few programs will meet his or her initial development and production plan [Biery, 1986:14].

2.2.4 *Relationships Between Cost, Schedule, and Performance Risk.* In some ways, risk management of innovative technologies is a zero-sum game. There will always be some intrinsic risk associated with novel development efforts that cannot be eradicated but can be portioned out between cost, duration, and the quality of performance for the project. This trading off may not happen in a quantifiable way, but is an often recognized risk management practice (e.g. expending more funds in an attempt to speed up development) [Klein, 1993].

Historically the majority of cost overruns in DoD weapon system procurement are due to schedule problems or technical difficulties, not underestimating costs. A recent study concluded that about 75% of cost growth in DoD programs was due to factors external to the program, such as unexpected changes in performance specifications, acquisition strategy changes, and budget difficulties. The rest were due to cost and schedule estimate errors and inadequately scoped engineering and software development efforts [Biery, et. al., 1994:75]. Schedule slippage is often the manifestation of technical problems, which then require greater than anticipated resources to complete [Biery, et. al., 1994:75].

The interrelationship of technical cost, schedule, and performance risks can be made clearer through careful analysis. This valuable understanding of the risks involved is what studies like this one try to bring to the decision maker.

2.3 *Dealing With Expert Judgement*

As RAND analyst E. S. Quade observed about 25 years ago, "Intuition and judgement permeate all analysis... As questions get broader, intuition and judgement must supplement quantitative analysis to an increasing extent" [quoted in Millett and Honton, 1991:43]. We must use expert judgement to judge the risks of emerging technology. Obtaining and quantifying input data is probably the most crucial part of performing risk assessments. It is a crucial but generally overlooked issue [Hudak, 1994:1025]. As such, it deserves detailed attention.

2.3.1 *Subjective vs. Objective Information.* Much of the input required in a risk assessment can only be found through information gathered from experts. In many cases this information will be very limited and may contain gross assumptions by an expert trying to bound the desired data with a lowest and highest conceivable value [Hudak, 1994:1026]. In assessing technical risks, analysts often find only one or two specialists sufficiently familiar with the program and technology to offer an assessment. These assessments are based on personal judgements [Biery, et. al., 1994:64].

Estimated probabilities are often used to build input distributions of random variables for simulation and other analyses, such as in this study. For our decision support model to be valid and accepted, it is important to understand common difficulties with subjective probability estimates of the sort used here. The choice of the family of distributions used is a crucial one.

Abstracting uncertainty with subjective probability distributions may or may not lead to better risk management, but such action often creates the illusion of doing so

[Troxler and Schillings, 93:230]. Care must be taken to avoid confusing these formalized expressions of uncertainty with statements of fact, especially with the decision makers. These subjective distributions are two steps away from the real world behavior being modeled — we are first saying that the future will be one of a set of potential outcomes, then we are estimating what the likelihood of those outcomes are (subjective uncertainty). Accurate objective data is always preferred, but when it is not available we must work with the best estimates we can get.

There is a danger when using experts of falling into the "expert halo" trap. It is easy to place undue credence on the opinions of experts. The analyst has the prestige of "expert" authority behind his or her study, while the uncritical decision maker is more likely to feel snug and secure under the protective umbrella of an impressive array of expert opinion. This tendency can make no one accountable, especially when estimates are made from group techniques such as the Delphi method. The analyst or decision maker can always claim that he or she was using the best advice possible and he or she is not responsible for what the experts say [Sackman, 1974:34]. While there are elements of truth to this, responsibility must still fall on the analyst.

2.3.2 *Quality of Expert Opinion.* Selecting experts to provide estimates is a problem in and of itself. Especially in cases of innovative technology, the set of potential sources of information may be quite limited. Chicken describes one way to discriminate between potential sources of expert estimates by quoting the methods advocated by the World Bank in selecting consultants [1994:177-8]. Adapting this method to our

requirements results in a subjective scoring scheme based on three criteria: a firm or individual's general experience with the technology in question, the proposed work plan for developing the estimate, and the qualifications of the key person(s). These three criteria are scored on a scale of 1 to 100 by the evaluator. The overall rating is obtained

$$\begin{aligned} S &= \sum_{i=1}^{3} w_i\, s_i \\ &= 0.15\, s_1 + 0.35\, s_2 + 0.5\, s_3 \end{aligned} \tag{2.3}$$

Adaption of the World Bank's Guidelines for Selecting Consultants

Criteria	Score (1-100)	Range of Weights w_i and Typical Value
Firm or Individual's General Experience	s_1	0.1 - 0.2 0.15
Work Plan	s_2	0.25 - 0.4 0.35
Personnel Qualifications	s_3	0.4 - 0.6 0.5

Table 2.1 [Chicken, 1994:177]

by a weighted sum of the three criteria, where the weights are determined by the evaluator based on his or her judgement of the criterion's significance. Table 2.1 describes the suggested weights. The resulting overall scores, using the typical criteria weights recommended by the World Bank, would then be:The higher the overall score, the better the subjective evaluation of that source of expert opinion. Note that the World

Bank's advised weights make the qualifications of the key personnel three and a third times as important as the firm's experience with the technology [Chicken, 1994:49-50].

2.3.2.1 *Training Experts to Provide Information.* One way to avoid biased estimates is to train the experts providing the estimates first. Guidelines and definitions can be worked out ahead of time in insure consistency across the range of experts. While this is an obvious suggestion, orientation and training is often overlooked [Biery, et. al., 1994:68]. Makridakis, et. al., note that even individuals who know a lot about the variable to be estimated may have trouble making subjective probability assessments, unless they are given guidance on how to proceed [1983:647].

2.3.3 *Soliciting Information From Experts.* There are many ways of gathering the opinions and assessments from the key people found to have the necessary special domain competence required for a technology forecasting study. The manner in which this information is gathered can have a large effect on the results, and so every effort should be made to make this communication process as clear and unbiased as possible. Little attention is often given to the critical step of acquiring expert judgement [Hudak, 1994:1025]. Therefore, we will discuss it in some depth.

2.3.3.1 *Interviews.* Interviews are a well-known and often practiced technique to gather information from experts. Virtually all corporations and analysts doing technology forecasting use interviews to gather information. The interview attempts to gain the in-depth judgement of the expert about the topic and goes beyond the more limited and structured form of written expert judgement found in a literature review.

Unless just one person is known or trusted to have all the information required to provide the forecast, conducting and synthesizing the results of numerous interviews is necessary [Millett and Honton, 1991:45-7].

There are several books and articles which give advice on planning and conducting interviews, but some basic practices taken from Millett and Honton are:

a) Plan the interview. The interviewer needs to give thought to whom should be interviewed and why. Interviews of experts should not be planned and conducted carelessly. The types of information needed should be identified first, then the names of people expected to supply it should be found. The number and extent of the interviews depends on the amount of time and funds available, balanced against the importance of the information. Questions should be written down in advance, to help capture the information the interviewer needs.

b) Conduct the interview in person or by telephone. Shorter interviews can be conducted by phone, but longer ones should be done in person. Face-to-face interviews have several advantages: the subject is more free to respond to questions in his or her own way, additional information in the form of facial expressions and body language can be gathered, and a personal rapport between interviewer and subject can be established. Phone interviews are less expensive in both time and funds, however.

c) Coordinate the interview with the subject in advance. The time and place of the interview should be agreed on beforehand. A letter explaining the purpose of the interview with perhaps sample questions should be sent in advance to the subject.

d) Always telephone when previously arranged and arrive for the interview on time. The interviewer is the supplicant — exhibiting bad manners is a poor research technique.

e) Ask questions in your own way and let the subject answer in his or her own way. Let the subject provide additional insight or information outside the formal structure of the planned interview. The interviewer must take care to listen to what the subject says, not what is expected. The interview should be a fair and realistic gathering of information, with the interviewer disturbing the results as little as possible [Millett and Honton, 1991:46-7].

The interview should be recorded in some way, either through taping or through detailed notes or transcription by the interviewer. If taped, care should be taken to inform the subject that he or she will be recorded. Their approval is required. This record should remain part of the project's documentation for later reference.

2.3.3.2 *Questionnaires*. Questionnaires are generally interviews prepared as written questions, to which the subjects reply without the presence of an interviewer. One can survey many more experts through questionnaires than through interviews. Many experts can be contacted at once, allowing a statistically large sample to be gathered where sufficient numbers of experts exist. The questionnaire can solicit information according to the specific structure required, in the terms and units specified to be compatible with the planned analysis. Responses from the subjects can be saved as part of the project documentation so that no information is lost [Millett and Honton, 1991:48-9].

A significant disadvantage is that the structured questions and answers keep subjects from saying exactly what they think. The structure limits the information that can be gathered to merely what was thought of during construction of the questions. One can get answers to what was asked, but there is no guarantee that the questions being asked are the right ones. Care must be taken that the writer of the survey and the respondent utilize the same definitions of terms used in the subject matter. Questionnaires can be misleading and confusing, and even irrelevant. Furthermore questionnaires are often costly and time consuming, as they require time and money to construct and refine, send out, and compile the answers. Of course, not all the recipients will respond — Millett and Honton suggest that a 75% return rate is excellent, and that even 25% can be acceptable [1991:48-9].

Constructing and executing questionnaires is a key task in survey research. There are a number of works on this topic (in particular, see Sudman and Bradburn, *Asking Questions: A Practical Guide to Questionnaire Design* (San Francisco: Jossey-Bass, 1982)). Millett and Honton suggest the following:

a) As with interviews, determine the kind of information required and why it is necessary before constructing the questionnaire. The purpose should guide the structure.

b) Select participants carefully to assure participation. While the ideal case would have all the participants and their specialties being known by the questionnaire builder, generally a proven mailing list of the kinds of needed experts is best used. The

group of recipients should have the necessary domain knowledge required for the estimates being sought.

c) Keep the questionnaire as short as possible. The shorter the questionnaire, the more likely the recipients will fully complete and return it. The questions should be focused on the goal and not be extraneous.

d) Structure the questionnaire, but leave the subjects the opportunity to express their own views. The questions should not solely be "true/false" or multiple choice. There should be essay-type questions that ask the subjects to use their own words. The questionnaire should include space for subjects to add their own questions and add other comments.

e) Make the questionnaire as user-friendly as possible. The structure and mechanics should be simple and concise [Millett and Honton, 1991:48-9].

2.3.3.3 *Delphi Method.* The Delphi method is undoubtedly one of the most commonly used technological forecasting methods [Makridakis, et. al., 1983:652; Sackman, 1974:3] and is one that many experts have some familiarity with. As such, it deserves special mention.

This approach was originally developed at RAND Corporation and is essentially a method of obtaining a consensus from a group of experts. As such, it is often used to generate a consensus forecast. The objective of the Delphi method is to obtain a reliable consensus of opinion while minimizing the undesirable aspects of group behavior. Its application requires a group willing to answer specific questions relating to new

technological processes. These experts do not meet to debate these questions, but instead are kept apart from each other to prevent them being influenced by social pressures or other aspects of group interaction. This is often done through correspondence, arranged by a coordinating moderator [Makridakis, et. al., 1983:652-4]. An iterative approach of questioning takes place, with successive rounds including results from the previous round showing the items on which there was a general consensus. Each iteration may be accompanied by selected feedback from the experts. The anonymity of the participants, use of statistical measures to describe the previous results, and the iterative polling with feedback are meant to produce authentic consensus and valid forecasts [Sackman, 1974:4].

The approach is meant to allow a spread of opinion that reflects the uncertainties underlying the specific technological issues under examination, while narrowing the inner 50% quartile range as much as possible without pressuring the experts so much that deviant opinions are not allowed. This is achieved by asking non-conforming experts to justify their positions [Makridakis, et. al., 1983:654].

Advantages of Delphi include low cost, versatility, ease of administration, minimal time and effort on the part of participants and moderators, and the simplicity, directness, and popularity of the method [Sackman, 1974:31].

Despite its prevalence, the Delphi method has several flaws. Many of the difficulties with the Delphi method or with any questionnaire result fundamentally from a problem of sampling. Despite generally small sample sizes, statistical analysis and testing

is often not done. Graphs of the inner quartile range are often the only way the results are presented to decision makers. The statistical representativeness and experimental rigor of Delphi studies has been called into question [Sackman, 1974:14, 28-9].

Using the central tendency of pooled opinion as the best estimate of expert opinion may not be the best ... Instead of the experts converging to a single consensus, studies using factor analysis have found subgroups of experts that cluster together with consistent opinions [Sackman, 1974:29].

A concise summary of the objections to the Delphi method was made by Weaver in 1972:

> At present Delphi forecasts come up short because there is little emphasis on the ground or arguments which might convince policy-makers of the forecasts' reasonableness. There are insufficient procedures to distinguish hope from likelihood. Delphi at present can render no rigorous distinction between reasonable judgement and mere guessing; nor does it clearly distinguish priority and value statements from rational arguments, nor feelings of confidence and desirability from statements of probability [quoted in Sackman, 1974:31].

One way to mitigate these criticisms is to avoid using the Delphi approach to make the forecasts themselves. A Delphi session can instead be used to create the inputs to other forecasting methods, applying Millett and Honton's advice about expert judgement [1991:61].

2.3.3.4 *Other Group Methods*. There are many other forecasting methods using groups of experts besides the Delphi approach. In general, the motivation is to build a better, more representative estimate than could be done individually.

One technique is called "idea generation," which is not precisely a technology forecasting method but serves as a way to generate input information for forecasting or planning. Idea generation is a somewhat more organized form of brainstorming, and is similar to what others call "focus groups," "idea groups," "creative sessions," and so on. It is bringing together a relatively small group of experts to generate thoughts on a defined problem for a stated goal. These goals include identifying: new applications for existing technologies or products, candidate technologies for a current need, issues and factors to be included in a larger forecasting method, and implications and candidate strategies from forecasting studies to be included in management planning. This method identifies ideas without evaluating them further [Millett and Honton, 1991:53-4].

The procedure for idea generation are to convene a group of eight to ten experts and brief them on the topic and the process to be used. The experts are allowed to interact through speaking or writing, while a moderator records ideas on large sheets of paper tacked to the walls of the meeting room for continuous review. The group interaction is terminated when the experts show signs of fatigue and/or the discussion starts to wind down. The experts then openly vote on the five to ten ideas they like best. This open voting allows for some consensus and group influence, although it is not required or forced. A written report documents the ideas and the results of the voting [Millett and Honton, 1991:54].

This method works best with a small group of creative experts who know and respect each other, discussing limited topics with little emotional or organizational politics

content. The experts must remain civil and not attack one another's ideas [Millett and Honton, 1991:54-5]. As a group interaction method, however, some of the same criticisms of the Delphi method apply.

Another group approach to expert opinion is the nominal group method, originating with Professors Delbecq and Van De Ven at the University of Wisconsin at Madison in the late 1960s and early 1970s. It has a more concrete structure, designed to handle situations where other group methods fail to be constructive: when argumentative and/or domineering people must be included, when people who do not know or like each other are involved, when managers and staff members are mixed together, when the topic is sensitive or controversial, or when organizational politics need to be managed carefully so the group exercise does not do more harm than good [Millett and Honton, 1991:55-6].

The nominal group technique can be used for the same purposes as idea generation, and can also be employed to generate criteria to evaluate or screen alternatives of a decision. The procedure for this technique includes a briefing of the experts on the topic and the method being used. Ideas are silently generated on paper by each expert before any discussion begins. Each expert then shares one idea from his or her list, going around the room in turn. This allows each individual an opportunity to share his or her ideas. Questions are allowed for clarification, but not debate or even comments on the virtue of the speaker's ideas. The moderator records these ideas on large sheets of paper mounted around the room, as in idea generation. The round robin of experts taking turns speaking lasts for a number of rounds or until a time limit is reached (three or fours turns

and a minimum of two hours is recommended). Once this has been reached, the ideas are reviewed and checked to see if any ideas can be consolidated to reduce redundancy. Ideas are only combined if no one objects. Each expert then votes privately on the best subset of ideas, ranking them according to some scoring scheme determined by the moderator. The voting results represent the amount of consensus on the "best" ideas. The moderator tabulates the votes immediately so that all the participants know the results before they leave. A follow-up report documents the procedure, list of ideas, and the results [Millett and Honton, 1991:56-7].

These group dynamics approaches offer a combination of creativity and group participation. They require an experienced and talented moderator who knows how to set the proper friendly and businesslike tone and manage the group of experts, and who must not seem biased to the participants. Preparation should be extensive, including the selection of participants and the preparation of invitations and instructions mailed ahead of time. The location of the meeting should be away from the normal workplace of the experts, free from telephones and other distractions. The experts must be selected carefully. Participants must have familiarity and experience with the topic, but do not have to be the preeminent experts on the subject matter. They must also be reliable, certain to show up and contribute according to the instructions given. Only about eight to twelve people should be included in one group session, although multiple sessions on the same topic can be held and later combined. In general, these group sessions should take

between a half to a full day. More than one day will result in the experts getting restless and contributing less meaningful ideas [Millett and Honton, 1991:58-9].

Millett and Honton strongly recommend that at least two separate group sessions should be conducted for forecasting purposes: one of in-house or "company" people, who provide microscopic expertise and a organizational "buy-in" to the subsequent results, and one of outside experts for a macroscopic perspective without in-house bias. These different groups will generate contrasting and illuminating results [Millett and Honton, 1991:58].

2.3.3.5 *Problems With Group Methods.* Open discussion between groups of experts involves interactive human behaviors. There are sometimes problems with these behaviors that can bias the resulting consensus estimates. Some of the group approaches mentioned above attempt to prevent some or all of these difficulties, but one cannot get the advantages of group estimates without potentially suffering from their pitfalls.

Some of these pitfalls include [taken from Meredith and Mantel, 1995:730]:

a) The Halo/Horn effect: A person's reputation (good or bad) or the respect (or lack thereof) in which a participant is held can influence the group's thinking.

b) Bandwagon effect: There is pressure to agree with the majority (indeed, this consensus is the objective in most group techniques).

c) Personality tyranny: A dominant personality forces the group to agree with his or her opinion.

d) Time pressure: Some people may rush their thinking and offer estimates without sufficient reflection, just as to not delay the group.

e) Limited communication: In large groups, not everyone may have the opportunity to provide input. The more aggressive or loudest group members may have an exaggerated effect on the group opinion (this is what the nominative group technique is meant to counter).

There is the fundamental issue of consensus estimates to be resolved, as well. The Delphi method as well as the other group techniques mentioned above rely on the claim that pooled expert opinion is more effective than individual judgement. Instead of combining independently generated individual opinions (such as described below in section 2.3.5), the process of feedback and interaction between the group participants creates highly correlated results as the group converges to conclusion. Social psychologists have known of powerful tendencies for individuals to conform to group opinion in relatively unstructured situations, particularly if they are not highly motivated. It is possible that the consensus formed through these group interaction methods is a product of this behavior, not mutual education and analysis [Sackman, 1974:45-7]. Still, whether the group interaction is highly structured as in the nominal group technique or as free-form as a staff or committee meeting, group forecasting is pervasive throughout program management and must be included as another tool for technology management.

2.3.4 *Probability Distributions for Use In Subjective Probability Estimates.* Many of the techniques used in risk analysis require input variables that represent

characteristics of the system being studied, whether that system is a release pathway for hazardous materials, a safety evaluation of highway routes for radioactive material transport, a model for total life-cycle cost of remediation activities, and so on. When data can be collected on these inputs, traditional ways can be used to specify the actual distribution of the values of the input over its range. The two techniques generally used are: using standard methods of statistical inference to "fit" a theoretical distribution form to the data, with parameters selected by goodness of fit; or by using values of the data themselves to define an empirical distribution [Law and Kelton, 1982:155-6].

But in assessing emerging technology, we do not have the opportunity to observe sufficient data for either of these methods in most cases. Choosing a distribution in the absence of data relies upon the subjective estimates of expert judgement. Through theory, past experience, or understanding of the limitations of predictions, some form of distribution is selected by the analyst or expert to represent the random variable. The ideal distributions for cost and schedule subjective probability estimates are unimodal, continuous, of finite range, and capable of taking a variety of shapes or degrees of skewness [Biery, et. al., 1994:69].

There are four commonly used distributions for expressing subjective uncertainty through expert opinion. The uniform, triangular, beta (and the specific PERT beta), and gamma distributions are all candidates, with their specific pros and cons. While the normal distribution is one with which most engineers are familiar, the infinite tails lead to problems with risk assessment and technology forecasting. Specifically, the infinite

negative tail creates the potential for negative costs or completion dates. It is not appropriate here.

The first step is to identify an interval of values that the random variable takes on, through asking the expert for their most pessimistic and most optimistic estimates. Let these interval endpoints be called a and b, where $a < b$. Once this has been done, other questions are asked as necessary to try as assess the other parameters of the assumed type of distribution [Law and Kelton, 1982:204-5].

2.3.4.1 *Uniform Distribution.* No other parameters need be estimated for the uniform distribution. Probability is evenly distributed between the two endpoints. Figure 2.8 shows a uniform distribution.

Uniform distributions are often used as a "first cut" at describing variables that are known to vary inside an interval but about which nothing else is known [Law and Kelton, 1982:158]. This is one way to transform the intervals described in section 2.1.2.2 for use in simulations.

Uniform Distribution Characteristics

Parameters	a, b	
Range	$[a, b]$	
Density	$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$	
Cumulative Distribution	$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$	
Mean	$\dfrac{a+b}{2}$	
Variance	$\dfrac{(b-a)^2}{12}$	
Mode	does not uniquely exist	

Table 2.2 [Law and Kelton, 1982:158-9]

Uniform Distribution Function

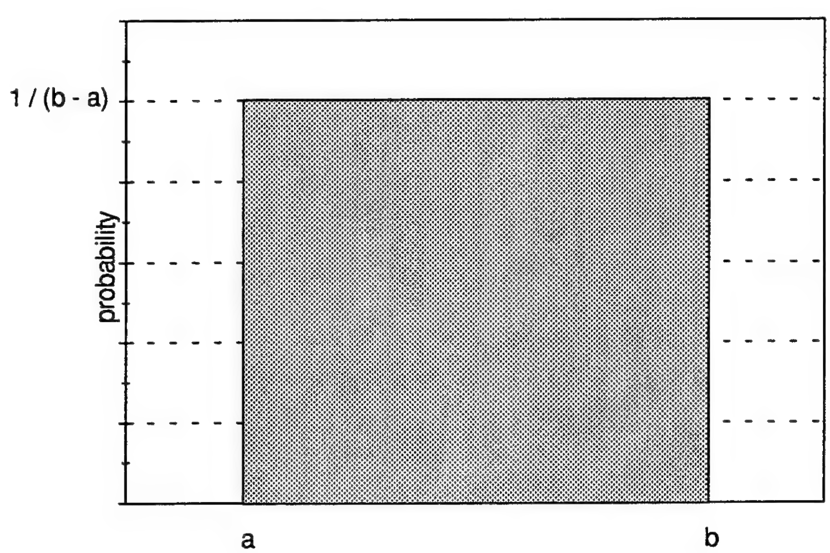


Figure 2.8

Triangular Distribution Function

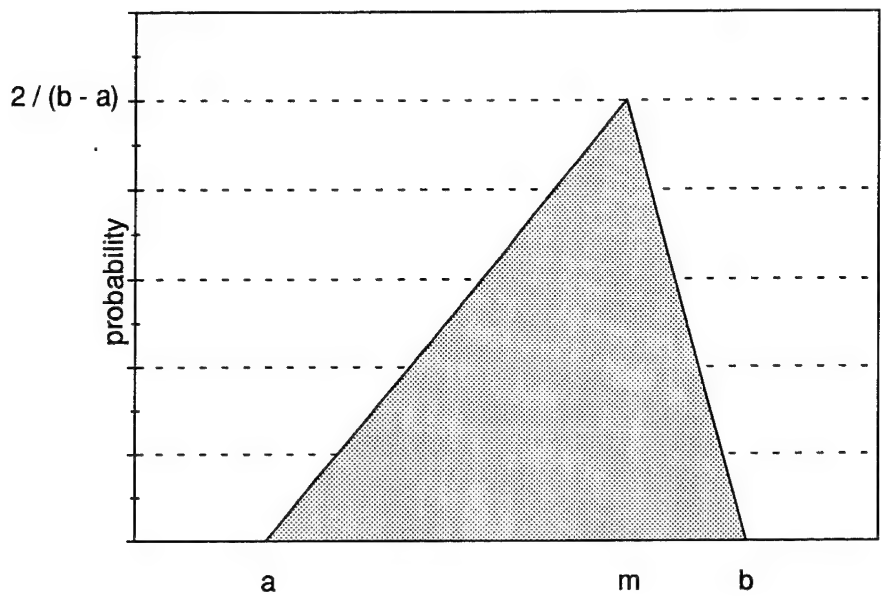


Figure 2.9

2.3.4.2 *Triangular Distribution.* The triangular distribution requires one other parameter to be fully specified, in addition to the interval endpoints. Experts are also asked to estimate the most likely value of the random variable, *m*. Armed with these three parameters, *a*, *m*, and *b*, a triangular distribution such as the one shown in Figure 2.9 can be used to represent the random variable of interest, *x*. Table 2.3 describes the mathematical characteristics of triangular distributions.

The triangular distribution is often used as a rough model in the absence of data [Law and Kelton, 1982:167].

The triangular distribution is easy to use mathematically and can take many unimodal shapes through changing the three parameters a, b, and m [Biery, et. al.,

1994:71]. If a = m or b = m, a right triangle is formed extending to the right or left, respectively [Law and Kelton, 1982:168].

Triangular Distribution Characteristics

Parameters	a, b, m
Range	$[a, b]$
Density	$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(m-a)} & a \le x \le m \\ \dfrac{2(b-x)}{(b-a)(b-m)} & m < x \le b \\ 0 & \text{elsewhere} \end{cases}$
Cumulative Distribution	$F(x) = \begin{cases} 0 & x < a \\ \dfrac{(x-a)^2}{(b-a)(m-a)} & a \le x \le m \\ 1 - \dfrac{(b-x)^2}{(b-a)(b-m)} & m < x \le b \\ 1 & x > b \end{cases}$
Mean	$\dfrac{a+b+m}{3}$
Variance	$\dfrac{a^2+b^2+m^2-ab-am-bm}{18}$
Mode	c

Table 2.3 [Law and Kelton, 1982:167-8]

2.3.4.3 *Beta Distribution.* The beta distribution requires two additional parameters to be specified, α and β. These parameters are not easily explained, as they interact to specify the shape of the distribution. This flexibility allows the beta distribution to taken on an infinite number of unimodal and bimodal shapes over the interval $[a, b]$ (the bimodal shapes are restricted to only those distributions with modes at the endpoints).

Figure 2.10 shows a typical unimodal beta distribution of the type often used in schedule and cost distributions.

A special case of the beta distribution that has been used for years in program management is the PERT beta, named for when it was first introduced for use with PERT charts. This technique uses the upper and lower limits together with the mode, *m*, to approximate a beta distribution's mean and variance [Keefer and Bodily, 1983:596]:

$$\text{PERT mean} \approx \frac{a + m + b}{6} \tag{2.3}$$

$$\text{PERT variance} \approx \left(\frac{b - a}{6}\right)^2 \tag{2.4}$$

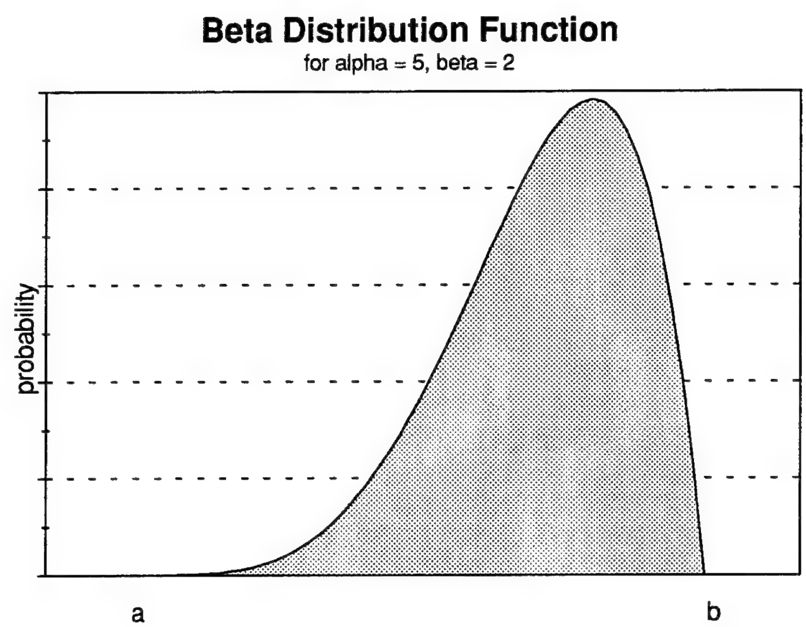


Figure 2.10

The PERT beta is a three point discrete approximation of an actual continuous beta distribution. Its accuracy in approximating the mean and variance is poor, especially when compared to other three point methods such as the extended Pearson-Tukey [Keefer and Bodily, 1983:601-2]. The original PERT assumption that the duration standard deviation is one sixth the range, generated from a general appreciation of project activities, has been discredited [Williams, 1992:266]. Because of its shortcomings and despite its previous popularity, we will not use the PERT approximations anywhere in this study.

Beta Distribution Characteristics

Parameters	a, b, α, β
Range	$[a, b]$
Density	$f(y) = \begin{cases} \dfrac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} & y = [a + (b-a)x],\ a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$ where $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
Cumulative Distribution	no closed form
Mean	$\dfrac{a\beta + \alpha b}{\alpha + \beta}$
Variance	$\dfrac{\alpha\beta(b-a)^2}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Mode	$\dfrac{\alpha - 1}{\alpha + \beta + 1}$ when $\alpha > 1,\ \beta > 1$

[Law and Kelton, 1982:167-8; Devor, 1987:163]

Table 2.4

2.3.4.4 *Gamma Distribution.* The gamma distribution is not bounded by an upper endpoint like the distributions mentioned above. Instead, it has an infinite tail. Two parameters are needed to fully specify a gamma distribution, α and β, where α is a shape parameter and β is a scale parameter. Since the range of a gamma distribution goes from 0 to infinity, one can represent a different lower limit by just starting the distribution at that point. Then a third parameter representing the lower limit is needed.

Gamma distributions are traditionally used with variables that have no upper limit, such as the time to accomplish some task [Law and Kelton, 1982:159].

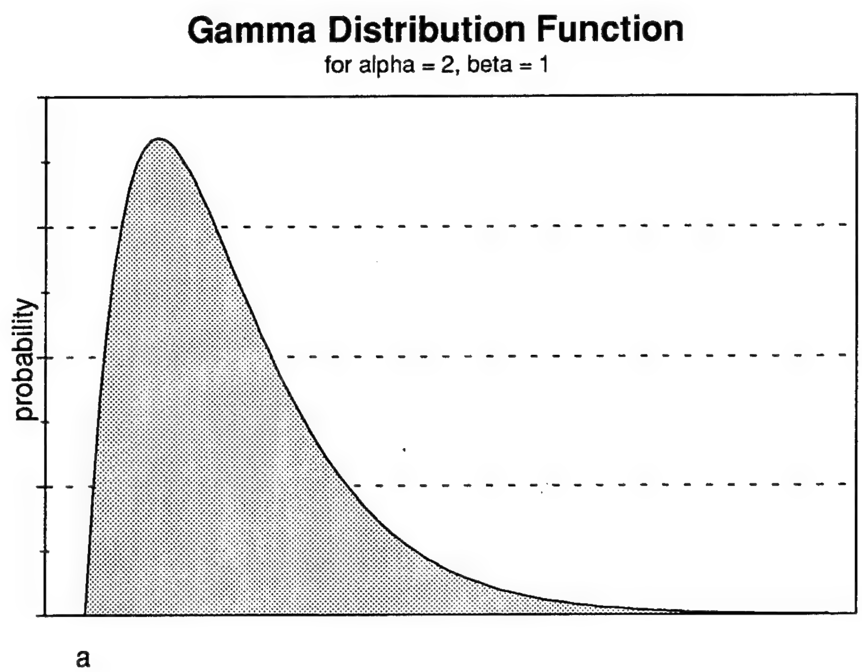


Figure 2.11

Gamma Distribution Characteristics

Parameters	a, α, β
Range	$[a, \infty)$
Density	$f(x) = \begin{cases} \dfrac{\beta^{-\alpha}(x-a)^{\alpha-1} e^{-\frac{(x-a)}{\beta}}}{\Gamma(\alpha)} & a < x \\ 0 & \text{elsewhere} \end{cases}$
Cumulative Distribution	$F(x) = \begin{cases} 1 - e^{-\frac{(x-a)}{\beta}} \sum_{j=0}^{\alpha-1} \dfrac{(\frac{x-a}{\beta})^j}{j!} & a < x \\ 0 & \text{elsewhere} \end{cases}$ when α is an integer, otherwise no closed form
Mean	$a + \alpha\beta$
Variance	$\alpha\beta^2$
Mode	$a + \beta(\alpha - 1)$ if $\alpha \geq 1$, a if $\alpha < 1$

Table 2.5 [Law and Kelton, 1982:159]

2.3.4.5 *Choosing A Family of Distributions*. The distribution used for representing input variables is an important choice when representing risk or uncertainty. The type of distribution becomes a framing question for soliciting information from experts about the random variable. Five criteria can be applied to help choose the type of distributions [from Williams, 1992:268]:

a) Easily understood: The parameters and assumptions involved with the distribution used must be easily understood by the expert providing the estimate.

b) Easily estimated: If the expert understands the nature of a parameter but finds its estimation to be unnatural, the quality of the estimate will be degraded.

c) Easily calculated: It is helpful if such information such as percentiles are easily calculated, letting an expert readily see the implications of choosing a particular parameter (corollary: this criteria suggests use of laptop computer and a plotting program be used to show the expert exactly what he or she is thinking of).

d) Limits: The ability to specify upper and lower bounds should be considered.

e) Particular Considerations: *A priori* assumptions, historical data, compatibility with other projects, and such need to be taken into consideration as well.

Recommendations from current literature are clear. The triangular distribution is the best compromise between simplicity, lack of knowledge, and ease of use by expert opinion. When the state of knowledge about a random variable does not even support the estimation of a most likely value, the uniform distribution should be used [Hershauer and Nabielsky, 1972:19; Law and Kelton, 1982:158; Haimes, et. al., 1994].

The triangular distribution is generally recommended over the beta for several practical reasons [Haimes, et. al., 1994; Williams, 1992; Biery, et. al., 1994]. Its simplicity and ease of use in simulations are strong motivators, as is the fact that only three parameters are necessary to completely define a triangular distribution while a beta distribution requires four (three for the PERT approximation). It is also easily estimated by experts. The beta, on the other hand, requires more information be known or assumed about the random variable in order to set the shape parameters. Betas are hard to solicit from experts, since these shape parameters are not intuitive satisfying. Experts unfamiliar

with probability find betas more difficult to understand [Williams, 1992:268]. A further disadvantage of the beta is that its use can artificially narrow the range of the random variable's distribution by implying a unjustified degree of precision. Smaller variances tend to result than with a triangular distribution for the same expert [Biery, et. al., 1994:71-2].

Where the imposition of a bound on one side of the distribution is unacceptable, the gamma distribution can be used [Williams, 1992:269]. While it also uses a non-intuitive shape parameter, the usefulness of the infinite tail may overcome this undesirable trait.

Other distributions than the four described here can of course be used. The choice should be made based on the characteristics of the random variable being estimated as well as on the simplicity, ease of use, and explicitness of the distribution. Care should be taken when employing normal and log-normal distributions, however. Systemic errors in estimation invalidate the central limit theorem. The presence of these kinds of errors makes the use of normal and log-normal distributions unjustified [Haimes, et. al., 1994].

2.3.5 *Using Subjective Probability Estimates*. Any information based on subjective assessment of the probability of future events is susceptible to bias. Some biases are obvious, while others are more subtle, difficult to perceive, and hard to deal with. The technical expert providing the subjective assessment may have a vested interest in the project in question, leading to some skepticism about the assessment's objectivity [Biery, et. al., 1994:64].

2.3.5.1 *Activity Duration Estimates.* Projects are made up of tasks that involve definite beginning and endings. They can be modeled through graphical displays called networks which are composed of activities and events, where activities show action or tasks to be accomplished and events show the completion or start of such activities. The network models the precedence relationships that exist between the various activities [Hershauer and Nabielsky, 1972:17].

Once the project network has been established, the next step is to estimate the duration of activities. The precedence relationships between activities can be used to determine the resulting duration of the whole project. Thus the estimates of the activity durations is critically important both in estimating the actual schedule of a project and in finding the expected "critical path," the interconnected activities that determine the overall project duration. If the activities on the critical path can be somehow shortened, the overall project schedule can be shortened as well.

For our purposes of examining schedule risks of new technology, we have only a few choices of ways to estimate these activity durations. If one feels certain about the length of time a task will take, based on historical evidence or past durations of similar activities, one can use a single point estimate to represent the necessary duration. This is the technique used in the Critical Path Method. Depending on the availability of historical data, probability distributions based on the frequency of past durations can be employed. If less is known, subjectively assessed random variables must be used to represent the time required for the task [Hershauer and Nabielsky, 1972:17-8].

Mistaken "Learning" Hypothesis

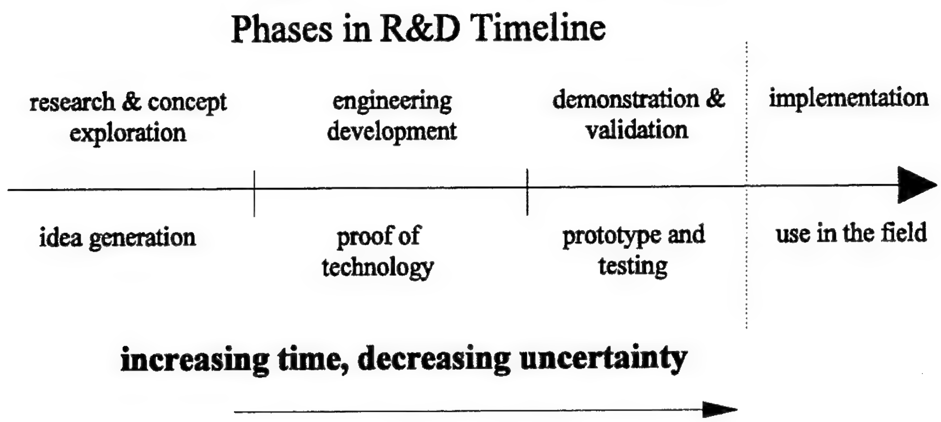


Figure 2.12

One would intuitively expect that estimates of project-related variables like schedule completion dates would get more accurate the closer one comes to the actual completion of the project, as shown in Figure 2.12.

Unfortunately, this is not the case. King and Wilson found that the accuracy of aerospace contractor estimates of the time remaining before contracted tasks were completed remained poor from long before the task began throughout the actual progress of the task. There was no improvement in accuracy until three weeks or less remained before actual completion. Their empirical study found that the contractors they examined underestimated the time required by about 30% before the project began and by about 21% during it. There were many more underestimates than overestimates in the historical data they studied [King and Wilson, 1967:310-5]. Their conclusions have been supported by later studies [King, et. al., 1967:84].

This shows the intuitively pleasing "learning" hypothesis, that activity duration estimates should improve as the activity progresses toward completion, may be invalid. Project milestones can be estimated on a projected schedule, but in general such dates will be underestimated.

2.3.5.2 *Other Types of Estimates*. While the previous section focused on activity duration estimates, similar inaccuracies have been found with other estimates of other uncertain quantities. Evidence gathered over the past two decades suggests that experts regularly neglect the full range of probability distributions when they attempt to estimate them. These subjective estimates provided by experts are subject to potential biases, especially for extreme estimates. This can be attributed to the way people assemble and process information to arrive at judgements. People reduce the complex task of processing all available information to the use of a limited set of rules and heuristics. This process of reducing information aids in making judgements in a highly complex world. This approach, however, tends to neglect information, especially regarding highly unlikely events. These rare events are, by definition, within the tails of distributions. For example, Hudak reports that cost estimates received by the Ballistic Missile Defense Office (BMDO) often under-represent the most unlikely outcomes by neglecting the tails of the cost distributions [Hudak, 1994:1026].

The potential for these kinds of errors in making subjective probability estimates should always be addressed when preparing to solicit such estimates from expert opinion.

2.3.5.3 *Adjusting Estimates.* Hudak describes a way to adjust for the under-representation bias using triangular distributions. By assuming the expert's estimated bounds are actually interior percentile points (fractiles), one can "correct" the distribution by applying a closed form equation to find the "true" bounds of the distribution that, with the unchanged mode, will completely specify the distribution [1994]. His approach is complicated and involves the solution of a four-degree polynomial (please see Appendix H for his method). Keefer and Bodily describe a similar way to get the limits of a triangular distribution, given the 10% and 90% fractiles together with the mode value, by solving two equations simultaneously [Keefer and Bodily, 1983:599]. Let x_{05} and x_{95} reflect the 5% and 95% fractiles, respectively. Using x_0, x_1, and x_m to represent the lower limit, upper limit, and mode of the distribution, one can solve for any two points given the others by:

$$\begin{aligned}(x_{05} - x_0)^2 &= 0.05\,(x_1 - x_0)(x_m - x_0)\\ (x_1 - x_{95})^2 &= 0.05\,(x_1 - x_0)(x_1 - x_m)\end{aligned} \qquad (2.5)$$

2.3.6 *Combining Estimates.* Since identifying the best model or most accurate expert is not possible *a priori*, considerable research has been focused on combining forecasts. In general, combining estimates made by multiple experts or sources of prediction seems to result in greater accuracy than just through relying on one single expert opinion [Makridakis and Winkler, 1983:987]. This is true for aggregating quantitative forecasts as well as more qualitative ones. The basic approach is to combine the different estimates of the n experts into an overall estimate $\hat{x}$ by assigning each estimate x_i a weight w_i:

$$\hat{x} = \sum_{i=0}^{n} w_i \, x_i , \tag{2.6}$$

where the weights sum to one ($\Sigma w_i = 1$). There are three basic approaches to choosing these weights: simple averaging, Bayesian combinations, and statistical methods using the correlation between errors.

2.3.6.1 *Simple Averages*. The use of simple averaging between multiple estimates has proven relatively robust and more accurate than more elaborate schemes in many applications. It is a very simple approach, that does not require information to be known about the accuracy of the individual estimates or the correlations between their errors. The theoretical justification for simple averaging is lacking, however [Gupta and Wilson, 1987:356-7].

With simple averaging, equation 2.6 reduces to the following:

$$\hat{x} = \frac{1}{n} \sum_{i=1}^{n} x_i . \tag{2.7}$$

A growing body of empirical research finds simple averages of expert opinion to be quite effective, and that only a small number of experts must be included to achieve most of the total improvement possible with a much larger set of experts [Ashton, 1986:405].

2.3.6.2 *Bayesian Approaches*. One problem with the simple average approach is that we know different experts and forecasting methods have different

accuracies for a given application. If we have some idea of what those differences are, it makes sense to try and incorporate that information into the method used to combine the different estimates. Bayesian approaches try to use as much of the information available to the decision maker as possible in setting the weights of Equation 2.6 to improve overall accuracy.

The subjective probability distribution provided by an expert is interpreted as the outcome of an experiment. While the expert sees this estimate as an expression of his or her state of information at the time of the estimate, the estimate itself is information or advice for analyst or decision maker to incorporate into his or her **own** state of knowledge. The problem of combining the estimates of several experts is then seen as an inference problem where Bayes' rule is applied to determine the posterior probability estimate [Morris, 1977:680].

Some idea of the accuracies of the experts is involved with Bayesian combinations. An expert must have his or her opinions calibrated, by comparing estimates to their true value to reflect the assessment performance he or she has established in the past, or by assessing the confidence of the analyst or decision maker in the judgement of the expert. These calibrations are used to modify the combination of estimates in ways that depend on the dependence between experts and the form of probability distributions being estimated [Morris, 1977:682-7].

In a sense, the expert's quality is assessed first using the past performance of the expert and then by the decision maker or analyst's perception of his or her accuracy. The

variance or range of the expert's estimate probability distribution is used as a measure of the expert's confidence in his or her own precision — the tighter the distribution, the more certain the expert. The basic concept of Bayesian combinations is that the analyst or decision maker who is combining the estimates uses his or her subjective judgement about the accuracy of the experts, together with preconceived "prior" personal assessment of the estimate itself, to produce a combined estimate [Morris, 1977:693; Winkler, 1981:481].

Bayesian combinations are very sensitive to dependence between experts [Winkler, 1981:487]. Modeling anything but independence between experts seriously complicates the joint calibration process [Morris, 1977:682]. Indeed, experts can be expected to produce somewhat dependent estimates, if only from common training or experience, or from working from the same data [Winkler, 1981:480].

Combining forecasts with weights determined from subjective probabilities of accuracy, reflecting a decision maker's confidence in the forecast, has some theoretical problems while seeming intuitively satisfying. A forecast of the type we are hoping to make is an inductive hypothesis on the true underlying stochastic process of the random variable we are trying to predict, not a prediction of a specific realizable event. We are really trying to divine the form of the random variable, and then make some statement about the value we expect it to take on. The subjective "probability that the true value is estimate *i*" means nothing if the random variable is continuous or nearly so [Bunn, 1974:158-9].

2.3.6.3 *Other Statistical Approaches.* Statistics are often used to attempt to maximize the accuracy of the aggregated forecast by assigning weights which account for the dependencies among the individual models or experts and their relative accuracies. If one knew the covariances of the different estimates being combined, one could always find a combined forecast with a smaller error variance than any individual forecast [Newbold and Granger, 1974:135].

Unfortunately, we don't know the values of the covariance matrix for the different estimates in our case of technology forecasting. Instead, weights are often determined from past performance of the experts in a variety of statistical ways [Newbold and Granger, 1974:136].

One additional wrinkle in using statistical methods to weight experts' estimates is an approach documented by Hogarth in 1978 in his article "A Note on Aggregating Human Opinions," which tries to prescribe the number of experts to aggregate the opinions of in order to maximize the accuracy of the aggregated estimate [quoted in Ashton, 1986]. By using analogies to test theory, he developed an analytical model that yields what he called "group validity" as a function of the number of experts, their mean "individual validity," and the mean intercorrelation between their judgements. The experts are rank ordering alternatives. The "individual validity" he uses is just the correlation between that expert's estimate and the actual value being estimated. "Group validity" is the correlation between the actual value and the simple average of the group of experts' individual estimates. His model makes group validity an increasing function of the number

of experts and their mean individual validity, and a decreasing function of the mean intercorrelation between the experts' estimates. This allows the ability to examine the results of adding the $(k + 1)^{th}$ expert to a set of k expert's aggregated estimates, and shows that the group validity of the new set of $(k + 1)$ experts will not necessarily increase simply be adding an expert whose individual validity is greater than the previous k expert's group validity. It may be necessary that the mean intercorrelation between the $(k + 1)$ experts be less than between the original k experts. His model provides the necessary conditions for the mean validity to improve with the addition of the $(k + 1)th$ expert, under certain conditions. For a small group of experts to have near maximum group validity, of about eight to twelve members, Hogarth argues that the mean intercorrelation must not be too low (approximately > 0.3) and/or mean individual validity must not exceed mean intercorrelation, with little statistical bias in the mean estimates. The limiting case, where $k = \infty$, is the ratio of the average individual validity divided by the square root of the mean intercorrelation between the experts' judgements [Ashton, 1986:405-7].

Ashton presents the results of an experiment testing these concepts with quarterly estimates of *TIME* magazine short-run advertising sales. He found that Hogarth's analytical model was effective in answering the "how many" and "which experts" questions to get the most accurate estimates. Ashton's empirical results showed that overall group validity did increase rapidly with additional experts added in, while the variance of the validity decreased rapidly as well. Of course, one must know the actual

value being estimated to use this technique, and it is only appropriate if the rank order of the alternatives is important and not the actual level of the estimates [1986:412-4].

2.3.6.4 *Summary of Combining Forecasts.* While the data-based approaches discussed above possess some desirable statistical properties, including low variance in the final aggregated estimate, their empirical performance has been disappointing. These approaches are often out-performed, in terms of accuracy, by the simple averaging method [Gupta and Wilton, 1987:358]. Ashton quotes Einhorn et. al. as saying standardized biases (bias $\cdot$ σ) of experts had to be about 0.70 or more before simple averages were outperformed by other realistic alternative weighting schemes [Ashton, 1986:407]. This unexpected result may be due to the large *a priori* data requirements for these methods. In practical applications, this data is not usually available, and so past history is often used to determine highly incorrect variance-covariances between the different estimates, which leads to erroneous weights [Gupta and Wilton, 1987:358].

The Bayesian approaches to combining experts' opinions require either past data or a decision maker's subjective assessment of expert accuracy to calibrate the opinions and set the weights of Equation 3.6. These methods become very complicated when dependence of experts are included and when the probability distributions being estimated are not normal. The actual weights are very sensitive to the degree of dependence [Winkler, 1981:487].

Using an average of forecasts is undoubtedly better than using a "wrong" model or expert. Therefore, unless an adequate theory exists to describe the forecasted technology

characteristics or strong evidence indicates a particular method is better than all the others, it is desirable to use multiple sources of forecasts and average their estimates [Makridakis and Winkler, 1983:995]. In cases of expert opinion, where the underlying "models" remain unknown, simple averages should be used [Kang, 1986:695].

2.4 *Public Feelings About Technology Risk*

One of the difficulties of environmental remediation is balancing the different perceptions of the problems of both the public and the government. Often the cost effectiveness, timeliness, and performance concerns that DOE considers are not the primary issues that are critical to members of the local community, environmental organizations, and other stakeholders.

The public whom the DOE deals with are often called "stakeholders," a term that the DOE defines as "individuals and groups in the public and private sectors who are interested in and/or affected by the Department of Energy's activities and decisions" [DOE, 1995c:20]. Stakeholders in environmental remediation cases generally identify themselves, and may be part of the following groups: the Environmental Protection Agency, the Department of Transportation, other federal agencies, Indian nations, state and local governments, elected officials, environmental groups, industry and professional organizations, organized labor, education groups, citizens' groups, and local community members [DOE, 1995c:20].

The primary concerns of local stakeholders center on public, worker, and environmental health [DOE, 1995c:21]. While analysis of the risks that each of the

candidate technologies pose to health and the environment are outside the bounds of this study, some reflection of expected public reaction to the employment of these technologies at DOE landfills is appropriate to provide to the decision makers of EM-50. Other major concerns include: the magnitude and severity of the health risks involved with the use of the technologies; how they affect the future use of the installations where the landfill are sited; the cost-effectiveness of the clean-up; involvement of stakeholders in the employment decision process; compliance with EPA and OSHA regulations, to include the evaluation of health and environmental risks; and the impact of transportation and storage of waste [DOE, 1995c:21].

In many cases stakeholders do not trust the Department of Energy to deal with their concerns. Criticisms of DOE health and environmental risk analyses characterize them as narrowly framed, based on little substantive data and depending on many assumptions. They do not address social or cultural values which are not amenable to quantification, such as equity, peace of mind, aesthetic, economic, community, future, and sentimental concerns [DOE, 1995c:21-2].

The implications of using a certain technology option may trigger irrational reactions in the public. The way people feel about the health and safety risks of many technologies do not reflect a logical and reasonable understanding of the actual probabilities and consequences of potential problems [Wheeler, 1993:1-3].

The contrast between the federal government on one hand and the dissenting stakeholders on the other is often seen as the conflict between "scientific rationality" and

"cultural emotion" by the press and members of the public. Arguments tend to be reduced to simplistic, dualistic terms. This springs in part from misunderstandings and suspicion of "Science" by many members of affected communities and environmental interest groups, but it is also created by the lack of trust in the government. This disposition towards an "us vs. them" conflict is aggravated by the media's tendency to dichotomize the news, which simplifies the situation as a battle between opposing sides where one side has to "win" [Coleman, 1995:74-5].

Managers evaluating the risks of new technologies must understand that some stakeholders will view "risk" in a different light. Analysts and decision makers use value judgements to assess the impacts of technological risks, but stakeholders may not agree with these trade-offs. Their opposition to certain remediation options should be examined when choosing the best technologies for use at landfills near their communities. Cultural beliefs are an important social complement to addressing environmental problems [Coleman, 1995:73-4], and dealing with stakeholder concerns is a necessary part of practical remediation execution.

III. *Methodology*

This chapter outlines the methods used to address the technical risk of innovative remediation technologies being developed by the Department of Energy for stabilizing and remediating landfill waste sites. Risk will be considered in both the inputs for the overall decision support system and the ultimate recommendations presented to the decision maker. This chapter will develop the methodology used to characterized technical risk in the decision support system and describe the demonstration of the model for the sponsor in DOE/EM-55. Ways to quantify and view the risks of recommended technology portfolios will be demonstrated.

3.1 *Landfill Stabilization Focus Area Technology Selection Project*

In 1994, three graduate students in the Air Force Institute of Technology's Department of Operational Sciences began work to help the DOE with its decisions concerning remediation technologies [White, et. al., 1995; Jackson, et. al., 1995]. Their research focused on comparing the total life-cycle costs of the alternative technologies for the Fernald Environmental Management Project near Cincinnati, Ohio. A spreadsheet-based life-cycle cost (LCC) model was developed using historical data where available and simulation results for a technology not yet fielded. They delivered a comparison between vitrification (MAWS process), ex situ cementation, and dry removal processes based on the requirements of each approach to remediate waste similar to that at the Fernald site

[Jackson, et. al., 1995:2-3]. One area that the Fernald/MAWS study did not examine explicitly was the issue of technical risk.

This research was extended in 1995, with an eventual plan to produce a decision support system tool that would compare many innovative and proven remediation technologies to be considered for use at various landfills using LCC and technical risk criteria. This tool was meant to be used by the staff of the DOE Landfill Stabilization Focus Area manager, Dr Jaffir Mohuidden, and so would examine the decision factors Dr Mohuidden considered most important. A contractor, MSE Technology Applications Inc., teamed with AFIT's Operational Sciences department, is on contract to complete this work as diagrammed in Figure 3.1. The effort includes two AFIT master's theses together

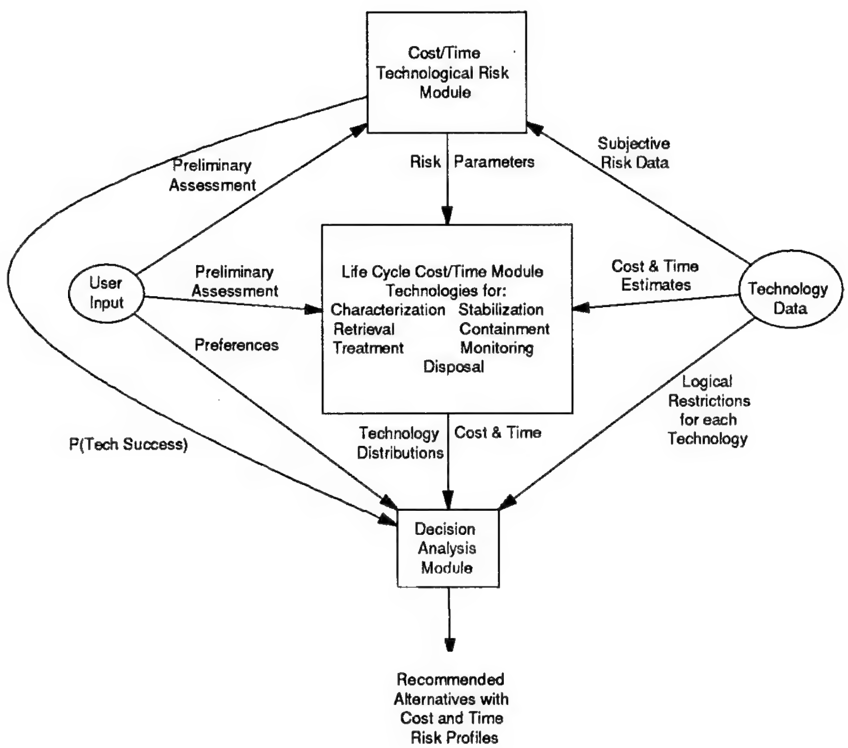


Figure 3.1

with a generalization and refinement of the LCC model from the Fernald/MAWS study by MSE employees.

The remediation technology decision support system includes "modules" for technical risk, life-cycle cost, and decision analysis. The structure and flow of information between the different modules is shown in Figure 3.1. The overall model will employ each of these modules, although not at the same time. Each will take information, act on it, and pass on a synthesis or judgement to the next. The penultimate synthesis is done in the Decision Analysis Module, which will compare alternative technology strategies according to criteria of cost and schedule, and will help the decision maker make better decisions about innovative remediation technology management.

The heart of the decision support system is the simulation of the remediation effort shown in Figure 1.2 as a network of sequential nodes that has a single path depending on choices made about stabilization and between retrieval-treatment-disposal vs. containment strategies. Each node represents the choice of one technology from a set of potential candidates. Each technology choice has a certain distribution of time and cost associated with it, drawn from expert judgement. State variables of the total time and cost are used to evaluate the performance of combinations of technologies. Draws from the chosen technologies' time and cost distributions are made as one moves from characterization through to monitoring. The sums of these technology costs and schedules make up the state variables for each simulation repetition, creating an overall distribution of time and cost over many repetitions for that specific combination or portfolio of technologies.

These distributions are then evaluated with utility functions for cost and time, which are combined in an additive multi-attribute utility function which is used to score the performance of each portfolio.

3.1.1 *Life-Cycle Cost Module.* The LCC Module is an outgrowth of the 1995 thesis work that simulated several competing treatment technologies applied to the Fernald site outside Cincinnati, Ohio. The 1995 models were very detailed, tailored for the specific technologies being compared at the Fernald site [Jackson, et. al., 1995:56]. The simulation that will be part of this study's overall model is less detailed but more flexible, to allow the comparison of many different technologies in up to seven different remediation processes. Less fidelity compared to the 1995 LCC modeling is the trade-off being made for the capability to simulate the remediation of any DOE landfill.

The LCC Module will produce probability distributions of operating cost and required processing time for each of the candidate technologies in each process in Figure 1.2. It will use expert opinion to estimate performance variables and cost elements as random variables, such as the cost per processing unit, the manpower required to operate such machinery, and so on. These input variables will feed into the LCC simulation from a database of technology information (see Figure 3.1). The simulation will produce realistic probability distributions for each individual candidate technology that account for correlations between real-world variables.

3.1.2 *Decision Analysis Module.* Once these probability distributions are generated for the different technologies, the Decision Analysis Module, using multi-

attribute utility theory, will develop the best combinations based on cost and schedule for the landfill. Net present value is used to discount costs back to the present day. Each of the processes from Figure 1.2 have technologies that are potential candidates for the best combinations. The DA model evaluates the overall schedule and cost results from employing these candidates in a total assembly of technologies called a "portfolio" or "technology strategy." Every potential combination of candidates is examined and its total cost and time distributions estimated. This information would then be available to the decision maker(s) when ultimate funding decisions are made.

Since the actual real-world decision to use a stabilization technique on a landfill is not made until after the characterization and assessment process is complete, using information about the waste stream that is currently unavailable, we cannot include it in our modeling. Adding a stabilization step to any technology portfolio adds additional costs and pushes the date of completion back. Since the DA model does not include environmental risk concerns that might motivate the use of stabilization, the added cost and time penalize the stabilization option so that it is never chosen. Because of this, the decision maker must decide *a priori* if he or she is evaluating portfolios including or excluding stabilization. Both cases could be run to see the effects of including it in the remediation strategy.

The operational costs and schedules of these candidate technologies are themselves random variables, with distributions resulting from the LCC Module. Therefore the DA model must account for the uncertainty in their performance.

Both cost and time are important to the decision maker. Unfortunately, there may not be a clear winning portfolio, with obviously better time *and* cost distributions. Multi-attribute utility theory is used to develop utility functions that allow the aggregation and trading off of cost and time in a way that reflects the decision maker's preferences. These preferences are used in the model to select the best portfolios. Interviews completed before the overall model is run establish these utility functions, which carry with them implied risk preferences as discussed in Chapter II. The relative importance of cost vs. time is represented by weights multiplied by each individual attribute's utility scores, which are then added together to get an overall utility for the aggregated cost and time of that portfolio. Absolute time and cost constraints are also used in the DA model to represent the limits of anticipated operating budgets or regulatory agreement deadlines. Instances of simulated remediations that have cost or schedule results beyond these constraints are assigned a total utility of zero. This effectively penalizes portfolios for sometimes exceeding these constraints, reducing the likelihood that it will be recommended.

3.1.3 *Technical Risk Characterization Framework.* The technical risk characterization framework consists of those processes that solicit and synthesize information specifically to allow the overall model to account for the technical risks involved with emerging, unproven technologies. As such, it consists of a set of procedures and recommendations requiring analyst judgement and discretion that cannot be completely automated.

The recommended decision strategies from the Decision Analysis Module are selected through picking those technologies that maximize the expected utility. The utility functions in the module include an indirect treatment of risk as explained in Chapter II, as they relate the decision maker's value to different schedule and funding estimates for the technologies. However, the explicit cost and schedule risks involved should also be presented to the decision maker, as expected utility may not provide all of the available and pertinent information.

The guidance received by the project team of AFIT/ENS and MSE emphasized that certain risks must be addressed in the modeling effort. Table 3.1 describes the specific major areas of concern.

Most of these risks lie in the "unknowable" section of Figure 1.2 at the point in time when the decisions must be made. They consist of events whose realization lies in

Risks Assessed in Technical Risk Characterization Framework

risk in...	method	used by
development schedule	distribution of dates when technology completes R&D	LCC Module
development costs	uniform cost per year of R&D	LCC Module
implementation performance	probability that technology will work successfully in the field	Decision Analysis Module
compliance with regulatory requirements	question user if the technology meets the regulation requirements governing the landfill in question	Technology Database (screening criteria)

Table 3.1

the future, but which must be predicted today. This is no easy task and requires technological forecasting methods to develop estimates.

3.2 *Sources of Information*

As already described in the introduction of this thesis, historical data is generally unavailable for use in forecasting the schedule, cost, and performance characteristics of the innovative remediation technology being examined in this study. As such, we are forced to rely on subjective judgements from those with specific domain knowledge about the technologies in question.

3.2.1 *The Developers of the Technologies*. Since the technologies in question are still in development or have recently been deployed, the pool of expertise available to produce detailed estimates of future capabilities, costs, and schedules is very small, and is primarily restricted to the contractors developing the technologies. Because of the level of detail required in the input performance variables and cost elements for the LCC Module, in-depth experience, both with the novel technologies being assessed and their development projects, is required to provide the necessary estimates. The luxury of selecting experts through scoring methods such as the World Bank's guidelines [Chicken, 1994:49-50] is not available to us because of the limited number of experienced people. This situation is problematic, as the principle investigators of a project may not be the objective, neutral judges one would prefer, nor are there other sources of information which could act as a check for potential bias.

The contractors developing these innovative technologies have a vested interest in remaining competitive. They must be optimistic about their progress to justify their continued work to their supervisors and DOE sponsors, as well as to motivate themselves toward quality performance. For these reasons, one must consider the possibility of unconscious biases influencing the estimates they provide for detailed schedule, cost, and performance-related analyses that influence future procurement decisions. Other conscious biases may exist as well, since they may well feel that future funding is somehow at stake. For these reasons, alternative sources of information and independent verification of technology developer estimates must be found when possible. Estimates and forecasts are biased and should be treated accordingly.

3.2.2 *Results from Similar Efforts*. Studies attempting to characterize the future capabilities and risks of remediation technologies have been published and can be drawn on to build the database of input variables for the decision support system(in addition to the technology developers). The Office of Technology Development produces summaries of the technology development projects funded under the different focus areas. The *FY-95 Technology Catalog: Technology Development for Buried Waste Remediation* and the *Landfill Stabilization Focus Area Technology Summary* provide overviews of the candidate technologies under consideration in this study [DOE, 1995a; DOE, 1995b]. While little specific programmatic or performance information is provided in these documents, the principle investigators and DOE contacts are listed. No characterization of risk is described.

Technical risks are described in a technical report completed for INEL on thermal treatment technologies [Feizollahi and Quapp, 1995]. Performance details and specifics are discussed. Unfortunately, these risks were only assessed qualitatively, using a low-medium-high scale [see pages 5-1, 5-41-3]. Some technology information for treatment techniques can be drawn from here.

A summary of remediation technologies was completed by a multi-organization committee on environmental technology that provides performance estimates for many of the candidates in this study [DoD, 1994]. The resolution of the operational cost and schedule estimates is not very fine for most of the technologies described.

3.2.3 *Combining Estimates*. As discussed in Chapter II, combinations of estimates from different forecasting methods and/or expert sources are often closer to the ultimate outcome than a single estimator alone [Makridakis and Winkler, 1983:987; Ashton, 1986:412].

For our problem of examining innovative technology, much of the information required for the more complex methods of weighting estimates does not exist. In most cases, we also do not have prior predictions from our experts that could be used to determine past accuracies. Until such records are kept by the Technology Development Office, the use of a simple average method is a reasonable choice for combining different estimates. Where the information needed for the inputs of the decision support system is provided by both the technology developers and published technology summaries such as mentioned above, they should be averaged together. Considering its performance in

comparison with many of the Bayesian and other statistical methods described in Chapter II, simple averaging may be the best choice where historical data would allow alternative weighting schemes [Makridakis and Winkler, 1983:987].

Averaging estimates from different people from the contractor may increase the accuracy of these forecasts, but they share the same potential biases and so their estimates could be highly correlated. This could actually lower the combined accuracy [Ashton, 1986:407].

3.3 *Procedures for Assessing Risks Through Model Inputs*

3.3.1 *Risks Involved With Regulatory Compliance.* The legal framework governing DOE environmental management activities is extraordinarily complex. The DOE must respond to the requirements of hundreds of permits, consent orders, and compliance agreements throughout dozens of legal jurisdictions at national, state, local, and tribal levels. Enforceable agreement milestones dictate the schedule of activities required by a permit or agreement. The compliance agreements are based on statutes which in turn evoke other statutes. These statutes are implemented through regulations, which in most cases include specific guidance on health and environmental risk [DOE, 1995c:11; see DOE, 1995d:H-1-6 for a listing of major laws and regulations]. Additional requirements may be levied by international standards such as ISO 14000 [Harmon, 1994].

The DOE has been negotiating agreements to address environmental violations at most of its major facilities since the mid-80s. Interagency agreements with the EPA and affected state governments have been reached for most of its sites on the National

Priorities List. Of the 117 agreements signed since 1989, 41 have been completed or renegotiated while 74 remain active [DOE, 1995c:15].

The DOE's remediation efforts are then driven by these legal agreements. A timeline and remediation standards for a given site are established in Records of Decision (ROD) that have the force of law [Mohuidden, 1995a]. Assessing the ability of the technical approaches to meet the remediation time and performance deadlines will be difficult to accomplish on a site-by-site basis. Unlike the other risk factors previously discussed, these requirements are known ahead of time and candidate technologies must be able to satisfy them (at least within the boundaries of our analyses). Therefore meeting this criterion is an absolute requirement for a technology to be considered for a given site.

3.3.1.1 *Procedure.* The complexity of the regulatory requirements makes a general examination of them problematic. These regulatory issues are best explored on a site-by-site basis because an examination of them in the aggregate is beyond the scope of this decision support system [Deckro, et. al., 1995].

Since the decision maker who is using the decision support system to help with his or her technology decisions will know which landfill is being considered, he or she is best suited to judge which, if any, technologies do not meet the regulatory requirements that cover that landfill. Therefore a simple series of screening questions prompting the model user to exclude those technologies that may not meet relevant regulatory requirements will be asked at the beginning of the DA module session. These responses, in conjunction with other site-specific characteristics, will reduce the set of potential candidate technologies

examined in the LCC and DA modules. Indicator variables in the technology database will be set that prevent excluded technologies from being considered for portfolios [Ralston, 1996].

3.3.2 *Schedule Risks in Research and Development.* The Department of Energy is planning for the long-term remediation of its landfills and other waste sites in the United States, but state and federal laws, in addition to other governmental agreements, place certain time restrictions on its actions. The DOE faces competing pressures to wait for lower cost remediation options to be developed and to begin clean-up operations immediately. Longer R&D schedules impacts the availability of potentially less expensive, faster, and safer remediation options in the field, and therefore the DOE would like to minimize these availability delays as much as possible. One of the overall purposes of this decision support system is to assist DOE technology managers in considering these trade-offs.

The DOE faces the possibility that a selected innovative technology will not be ready at its expected availability date. The planned use of such a delayed technology at a waste site could cause that site remediation effort to fail to meet mandatory deadlines. There is no guarantee that an ambitious technological approach will be successful — one estimate of the likelihood of technical completion for commercial R&D projects is only 60% [Bhat, 1991:262]. Other, more costly methods may have to be employed when the EM-30 or EM-40 manager becomes aware that a technology will not be available. In the face of such an outcome, the credibility of DOE's management of the nation's remediation

program would suffer. In terms of our risk definition in Chapter II, the negative consequences of schedule overruns could be very grave. The probabilities of these overruns must be estimated to have a complete picture of the risk involved.

3.3.2.1 *Procedure.* The availability of candidate technologies is estimated using a probability distribution of dates when the technology completes R&D (see Figures 3.2-3.4). This "release date" is defined as when the given technology has satisfied all of its specified laboratory and test performance criteria and is considered ready for use in the field. "Successful development" is therefore considered to be the point when the technology has met whatever test and demonstration standards that mark the final stage of R&D. In this fashion a technology in the early "idea exploration" phases will have a range of release dates that extends far into the future, while one that is very close to full development will have a range that ends in the near term (note that this approach assumes that, given sufficient (perhaps infinite) time and money, any technology will be successfully developed).

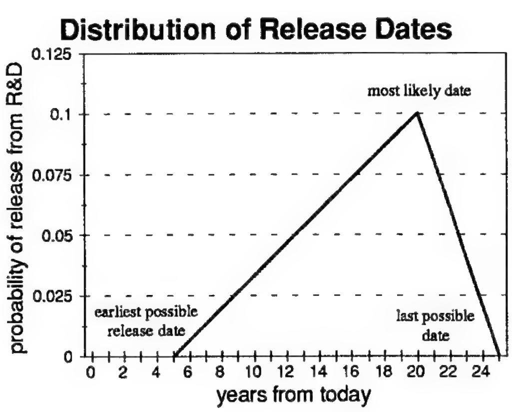


Figure 3.2

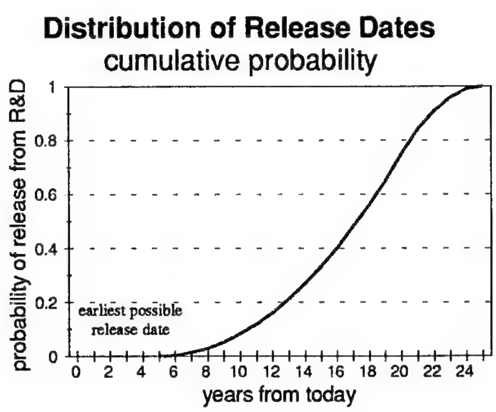


Figure 3.3

DOE Technology Maturation Phases

Basic Research	Applied Research	Exploratory Development	Advanced Development	Engineering Development	Demonstration	Implementation
Idea Generation		Proof of Technology		Engineering Prototype	Production Prototype	Utilization by Customer

progress of time

Figure 3.4

These "release dates" are estimated using a triangular probability distribution. Triangular distributions are a better choice than other distributions, such as the beta, for several practical reasons. They are easy for experts to estimate, requiring only three easily understood parameters. They are simple to calculate and understand, and can take on a variety of skewness shapes while being bounded by upper and lower limits (see Chapter II, section 2.3.4.5). The triangular distribution is available as a feature in a number of simulation codes. In the absence of other information that would allow the more precise determination of the shape of the release date distributions, the conservative assumption that the distribution is triangular will be used in this study [Biery, et. al., 1994:72]. The experts are asked to provide estimates of the release date for their technology based on a best, worst, and most likely case. This expert group of contractors developing the technologies has the best understanding of the technological breakthroughs, available resources, potential funding fluctuations, and other factors which influence the final completion date. If other expert evaluators are available, they can supplement or replace these contractor estimates. The resulting estimates, the earliest, most likely, and latest

R&D release dates, are used to define a triangular distribution of potential completion dates that the LCC model uses to establish an earliest possible implementation date.

3.3.2.2 *Adjusting the Release Date Distributions.* Examinations of the literature demonstrate that contractors generally underestimate the actual time required to accomplish tasks, and that such estimates remain inaccurate from before the task begins until a few weeks prior to completion, regardless of the actual duration [King and Wilson, 1967]. The tails of subjective probability distributions for activity durations (i.e. very short or very long) are also generally neglected [Hudak, 1994].

These potential errors and biases motivate the application of a correction to the contractor estimates. A wholesale adjustment to the estimated release date distribution should be done only if historical data exists that shows significant, consistent over- or under-estimation of completion dates by that expert. Without such empirical data, correction factors should not be applied to the mode date estimates. However, general adjustments to the tails of the release date distributions is supported by the literature. The Ballistic Missile Defense Office (BMDO) of the Department of Defense has been applying corrections to such contractor estimated probability distributions as standard practice [Hudak, 94]. Since predictions of the near future are generally more accurate than more distant predictions, a smaller adjustment factor is used for the earliest release date than for the latest release date. This conservative approach will help reduce the risks of seriously underestimating the actual development time.

The adjustment will follow a similar development as the bias-removal technique in Hudak [94]. Hudak provides a method to convert between the absolute bounds of a given triangular distribution and the inner fractiles using similar triangles that requires the solution of a complicated fourth degree polynomial, as already described in Chapter II. He recommends using 10% and 90% fractiles for the contractor-supplied estimates, as is done at BMDO. We will use 3% instead of 10% for the earliest release date, however, as discussed above (see Figure 3.5). The contractors' estimated earliest possible release date will be taken to actually represent the 3% fractile of the release date distribution. The estimate of the latest release date will be used as the 90% fractile. The new bounds are pushed outward, extending the range of the distribution.

Keefer and Bodily mention a simpler procedure to convert between fractiles and the bounds which will be used here [1983:599]. Extending their method to 3% and 90% fractiles, we can find the new earliest and latest release dates by solving the following equations simultaneously:

$$\begin{aligned} (x_{03} - x_0)^2 &= 0.03\,(x_1 - x_0)(x_m - x_0) \\ (x_1 - x_{90})^2 &= 0.10\,(x_1 - x_0)(x_1 - x_m), \end{aligned} \tag{3.1}$$

where x_{03} is the 3% fractile, x_{90} is the 90% fractile, x_m is the mode, and x_0 and x_1 are the lower and upper limits of the adjusted distribution, respectively. The solution to these equations involves a fourth degree polynomial, resulting in four potential solutions for x_0 and x_1. After excluding those infeasible pairs where one or both values fall inside the 3% and 90% fractiles, the remaining pair is the new lower and upper limits, respectively.

Solving the two simultaneous equations can be done using mathematical software such as *MathCad*© or *Mathematica*©, or by using numerical solution algorithms that exist for all major programming languages such as FORTRAN or C++ (see *Numerical Recipes* for an example).

Figures 3.5 and 3.6 show an example of applying this method to the release date distribution of one characterization and assessment technology, going from a triangular distribution based on an earliest date of 1, a mode of 2, and a latest of 4 years from now to one with an earliest date of 0.549, a mode of 2, and a latest of 6.330 years from now.

This approach is simpler than the one Hudak describes, which involves much more complicated algebra (see Appendix H). Tests of Hudak's method against the approach just described show that they are equivalent.

3.3.3 *Cost Risks in Research and Development.* Total life-cycle cost is EM-50's dominant criteria for selecting remediation technology, subject to the constraints of public safety and regulatory requirements [Mohuidden, 1995a]. The cost to develop a technology is an important part of that total remediation price tag. The risks here are that the actual development costs are larger than the DOE managers have predicted and funded. Should a development cost overrun occur that exceeds the contingency fund reserves in the EM budget, funding adjustments would disrupt the progress of other development projects as funds are shifted between projects. Such reallocations can

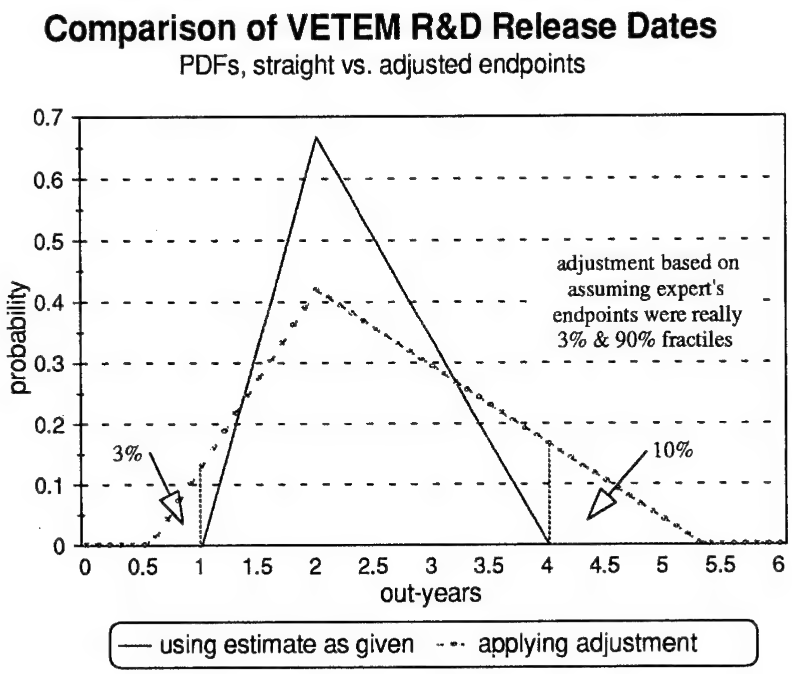


Figure 3.5

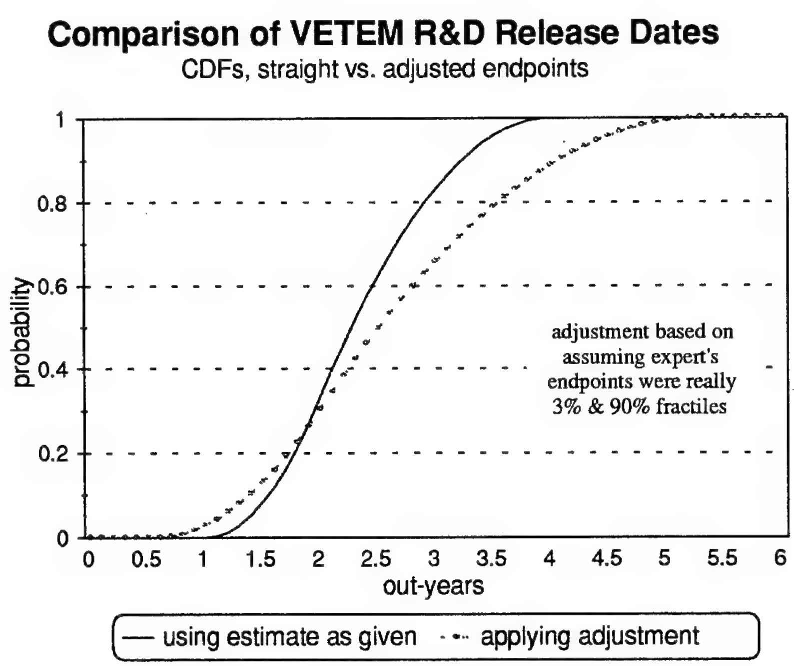


Figure 3.6

affect other projects' development schedules and ultimate deployment. The troubled technology's R&D may be stretched out and delayed due to insufficient funds, similarly affecting the final delivery date of the finished product. If the projected cost overrun is sufficiently large, the technology development may be cancelled altogether.

Accurately predicting the final development cost, however, is not easy, especially if long-term budget predictions from contractor proposals are not available. There are many factors involved in R&D costing, including time-dependent costs such as work force levels, capital costs such as laboratory equipment and prototype materials, organizational overhead and other related expenses. The final development cost for a program can be a function of what could be hundreds of individual random variables. However, the data needed to construct such a detailed cost function are unknown during the early stages of a project, and arguably are unknowable. While there surely are time-cost trade-offs that can be made, determining the actual relationship between schedule acceleration-deceleration and final cost is not empirically easy or theoretically certain [Biery, et. al., 1994:80].

The distribution of development cash flows over the R&D phase of a technology development project could conceivably take many shapes. The actual costs for a given year may be as dependent on programmatic factors outside the project, such as the availability of funds, as any technology-specific cost of development. In a multi-year, high visibility program like the DOE's remediation research efforts, there is a high likelihood of

budget fluctuations, both of less and more funding. The availability of funds is considered an issue outside the bounds of this study.

Since the products under development for this study are emerging technologies that extend the state-of-the-art in environmental remediation, there are further difficulties in predicting the final development costs. The progress of the development effort relies on innovative solutions to difficult engineering problems. The timing of these technological breakthroughs is impossible to anticipate, short of wizardry, as they are dependent on individual creativity, organizational action, and luck. While it may be possible to model the occurrence of these breakthroughs as some random process based on empirical research in other fields, the soundness of such a model will be impossible to validate using normally available (or rather unavailable) DOE technology development data.

3.3.3.1 *Procedure.* We know the development costs are strongly related to the time required to complete R&D. Workforce and O&M costs are directly dependent on the duration of R&D, while the costs of capital goods such as scientific equipment and engineering materials are not (this assumes that capital goods purchasing schedules are not materially affected by downstream delays over the length of the development program). Following Biery, et. al., we will assume that, in the absence of more precise data, all costs are linearly related to the actual time required to complete development [Biery, et. al., 1994:80]. Using the projected remaining development costs and development schedule gathered from the technology developers, a cost per unit time will be assigned to the

project that will be used in conjunction with the release date distribution in the LCC model to estimate the final remaining development cost. This cost is expressed as:

$$\textit{development cost per year} = \frac{\textit{projected remaining R\&D cost}}{\textit{median release date} - \textit{present date}}. \qquad (3.2)$$

This R&D cost per year will be stored in the Technology Database, where it will be used by the LCC model to calculate the final development cost. One run of the LCC simulation will yield:

$$\begin{aligned}\textit{total development cost} = \; & \textit{triang}\,[\textit{earliest}, \textit{median}, \textit{latest}\,] \\ & \times \textit{R\&D cost per year}. \end{aligned} \qquad (3.3)$$

3.3.4 *Performance Risks in Implementation.* The transfer from successful development to successful implementation is a step whose importance should not be underestimated. Even if a technology has passed all of its developmental test and evaluation (DT&E) requirements, there is still no guarantee that it will move satisfactorily to the field. DT&E rarely duplicates real-world conditions. Often the situations where the technology is put to use are different from those anticipated by the original technology developers [Leonard-Barton, 1987]. To account for these possibilities, one may be able to estimate the likelihood that a remediation technology is successful in the field after it was successfully developed in R&D.

Most of the overall decision support model focuses on the implementation of the remediation technology. The DA Module uses the R&D release dates and development costs as starting points for the distribution of costs and schedule milestones resulting from

the LCC simulation. Both the DA and the LCC modules assume that the technologies perform within the bounds set by the performance variables established by expert opinion — that is, the technologies will only act as well or as poorly as anticipated by the technology developers. The possibility of a technology failing to meet the expected performance criteria and requiring replacement by another technology to accomplish the remediation of the landfill must be addressed. DOE technology selection studies have used similar criteria [Feizollahi and Quapp, 1995:5-1].

The likelihood of implementation success depends on many factors; some are site dependent, others are driven by the technology, and by their very nature are unknowable until failure occurs. The question of a successful implementation must address the chance that the preliminary site assessment was incorrect. A mis-assessed site could contain other waste types and items which the chosen technology may not handle.

3.3.4.1 *Procedure.* This unknown implementation success will be modeled through expert opinion. The probability of implementation success is defined as the likelihood that the technology performs within expected parameters, with the understanding that the preliminary characterization of the landfill may not be correct, given that it was released from research and development. Let P(use) be the probability of successful use:

$$\begin{aligned} P(use) = P(&technology\ \ performs\ \ within\ \ expected\ \ parameters\ \ in \\ &field\ \ use \mid technology\ \ was\ \ released\ \ from\ \ R\&D\ \ and \\ &preliminary\ \ site\ \ assessment\ \ may\ \ not\ \ be\ \ correct) \end{aligned} \tag{3.4}$$

By making P(use) conditional on the technology being first successfully developed, we can consider the probabilities of successful development and successful implementation as being independent. P(use) is the likelihood that the technology works as planned once it has completed R&D. By accepting the assumption that the test and demonstration standards which a "successfully developed" technology must meet remain essentially unchanged through its multi-year R&D, we may assume that its P(use) is then independent of either the time or cost required for development. This assumption of independence is central to how we structure the overall model, as it allows us to consider development and implementation separately.

Without specific knowledge of the covariance of the cost and schedule effects of all the combinations of possible technologies, this assumption is required to accomplish any modeling at all. Again, the need for robustness is balanced against the decision support model's fidelity. Like democracy, this may be the worst choice for modeling a spectrum of landfill remediation technologies — except for all the others.

Obviously the likelihood of using a technology successfully at a site depends on the waste being in a form that the technology is capable of processing. For example, a treatment technology that cannot handle volatile organic compounds (VOCs) will not work successfully on a waste stream that unexpectedly contains VOCs. Given the state of uncertainty about the contents of DOE landfills across the country [Mohuidden, 1995a], we cannot guarantee that a technology will always face the kinds of waste material that it was designed to manage. Even with an acceptable characterization, a key hazardous

element could be missed in a site until remediation commences. Therefore, we have used expert judgement of the robustness of the remediation technologies, expressed through P(use) estimates, as a method of dealing with this possibility.

The Decision Analysis Module will use this probability as the controlling factor as to whether the technology works, adding its individual processing time and duration to the overall master schedule and costs, or fails, requiring a replacement technique that incurs additional cost and time to complete that remediation process.

3.4 *Assessing Risks of Recommended Alternatives*

There is one last crucial step in building risk assessment into the decision support model, so that the results of the model reflect the technical risks involved. The decision maker must have information on the relative riskiness of his or her decision alternatives available when making choices. A quantitative measure of risk must incorporate both the probability of undesired events and their consequences, and allow a decision maker to unambiguously distinguish between different alternatives using risk as a criteria. There are several ways to capture some estimate of risk for the decision maker described in Chapter II, including the mean and variance of the anticipated costs and scheduled milestone dates, the Jia-Dyer "standard measure of risk," and others. Since we have decided to express risk through the tangible attributes of cost and time, we will compare decision alternatives by comparing the estimated costs and schedules that result from the overall model.

3.4.1 *Histograms*. A convenient way to compare alternatives is to examine the results from the DA Module expressed in the form of histograms. These represent the

frequency of occurrence (probability distribution) of particular time and cost values for a particular portfolio. The fraction of occurrences where total costs or required time are intolerably high is obvious to the decision maker. All the information needed to express risk (the magnitude of the cost or time and the probability of occurrence) is available from the probability distribution functions (PDFs). However, such information is not presented in a concise, compact way. Comparing many alternatives requires examining many histograms. Alternative methods of expressing risk include ways of condensing the histogram's information in other forms.

3.4.1.1 *Getting Histograms From* DPL©. The DA Module is based in a *DPL©* model. After the model is run, the results are presented through a combination of windows including a distribution window that displays the cumulative probability distribution of the attribute selected in setting up the run (cost, time, or total utility). Clicking on the "graph" menu in that window presents the option of viewing the "cumulative" distribution (the default), a "frequency histogram," or a "frequency X-Y" graph (an alternative form of the frequency distribution). Selecting the frequency histogram will result in a graph similar to Figure 3.7.

Obtaining the information contained in the histogram is accomplished by using the options under the "file" menu. These save the histogram in a text file that can be imported

into a spreadsheet with little difficulty. One can choose to "export as displayed," which creates a file allowing the reconstruction of the histogram graph, or to "export interval midpoints" of the histogram bars for later analysis.

3.4.2 *Classic Utility Theory.* As mentioned in Chapter II, classic utility theory as established by von Neumann and Morgenstern [1947] includes an indirect way to express the decision maker's preferences toward uncertain outcomes. The Decision Analysis Module uses utility functions to characterize the relative values of total cost and total time required to remediate a landfill in selecting the best technology portfolios for the given remediation task.

The shape of the utility function and the local risk aversion, $-\frac{u''(x)}{u'(x)}$, can be

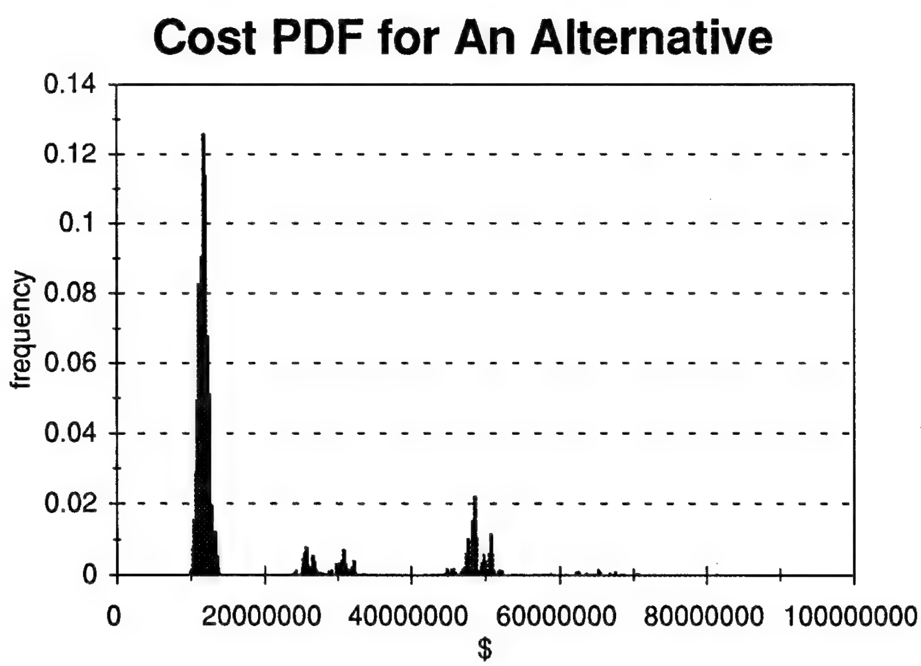


Figure 3.7

examined to understand the decision maker's preferences for risk. There is, however, some difficulty in interpreting these indications of risk preference if the utility function is complex.

3.4.2.1 *Risk and the Utility of an Alternative.* In our technology management decision, we prefer less cost and shorter schedules to more cost or longer schedules. Therefore we consider only decreasing utility functions. The utility function $u(x)$, assessed for the attribute x, expresses the decision maker's value for different levels of x. When x is the expected outcome of a risky decision, expressed through a reference lottery, the shape of the utility function expresses the decision maker's risk attitudes [Keeney and Raiffa, 76:180].

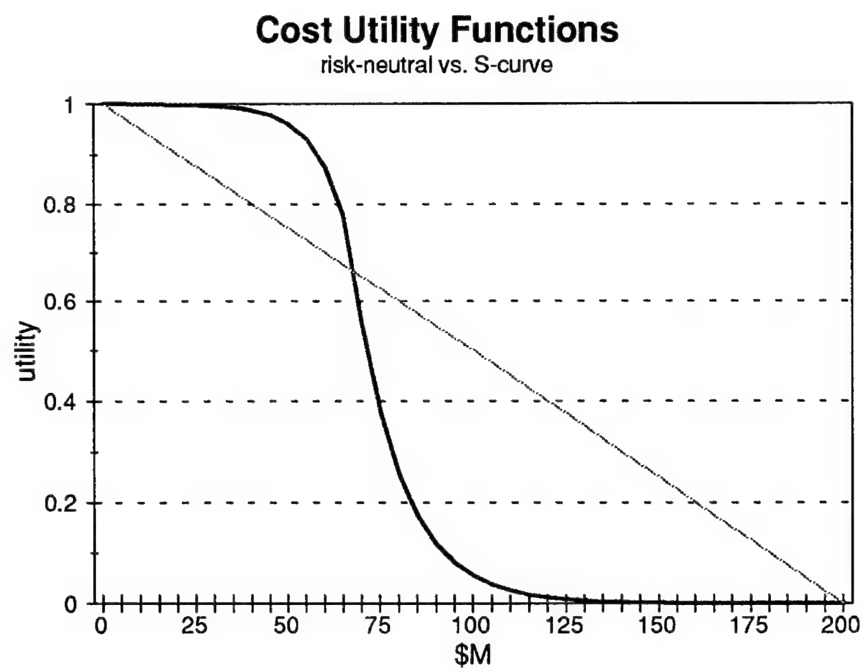


Figure 3.8

Consider the utility curve for remediation costs used in the DA model in Figure 3.8, shown compared to a risk neutral utility function. Examining the shape of this S-curve suggests that it is risk averse from 0 to about \$65M, and risk prone beyond \$65M. That is where the second derivative of the S-curve utility function changes sign, and therefore where the local risk aversion function goes from positive to negative.

To examine the way risk can be measured through this utility function, consider two different hypothetical alternatives, #1 and #2. The cost frequency distributions are shown in Figures 3.9 and 3.10, respectively. Clearly alternative #2 exhibits more variance than alternative #1. The mean cost of #1 is \$65M while the mean cost of #2 is \$51M.

We can apply the S-curve utility function from Figure 3.8 to these alternatives and obtain the results shown in Table 3.2.

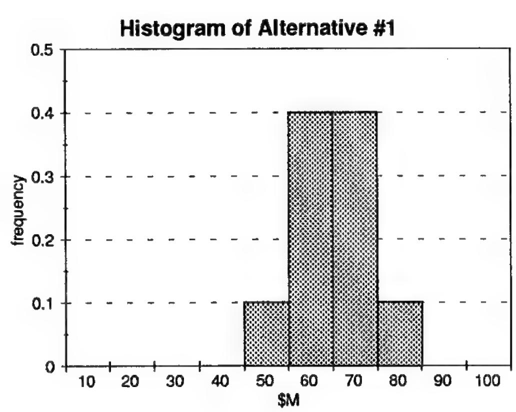


Figure 3.9

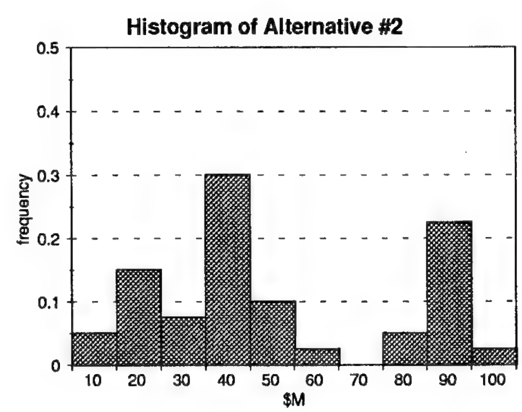


Figure 3.10

Comparison of Two Example Cost Alternatives

	Alternative #1	Alternative #2
Mean ($M)	65	51
Expected Utility	0.692	0.729
Certainty Equivalent ($M)	67.05	66.37
Risk Premium ($M)	2.05	15.37

Table 3.2

Alternative #2 has the higher utility and so would be ranked higher than #1. It has the lower certainty equivalent (CE). If one looks at the difference between the CEs and the means, the risk premium, one can see that #2 has a much higher risk premium. This represents how much the decision maker would be willing to pay for another alternative that would have no uncertainty involved with the remediation cost. The risk premium is therefore an indirect measure of the risk associated with #2's cost distribution.

An equivalent way to look at these alternatives is to develop PDFs of the cost utilities for these technology alternatives, resulting from the application of the utility function to the cost PDFs. These utility PDFs are shown on Figures 3.11 and 3.12. The means of these utility PDFs are 0.692 and 0.729, consistent with the expected utilities of the cost distributions. The decreasing utility function of Figure 3.8 can be thought of as a non-linear transformation of the cost PDFs, where the general shape of the cost PDF is preserved but reversed. Because of the S-curve shape of the utility function, more weight is preferentially given to the smaller costs than the larger ones. This “spreads out” the shape of the original cost distributions.

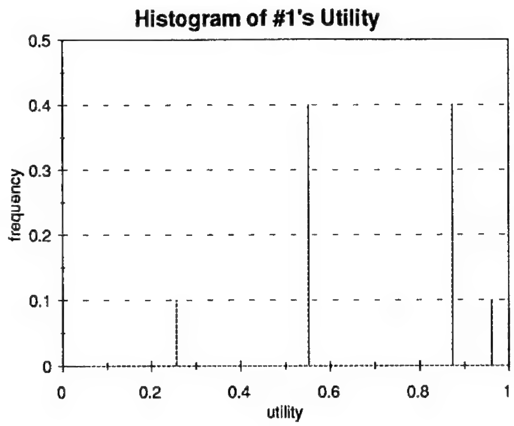


Figure 3.11

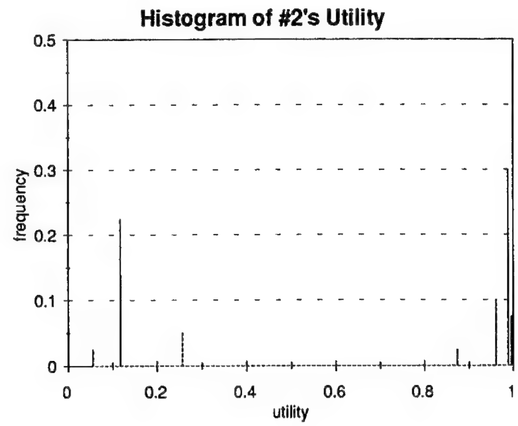


Figure 3.12

The difference in shape between the cost and utility PDFs is due to the utility function, and therefore the shape difference shows the "risk preferences" of the decision maker (*assuming* the utility function has been correctly assessed and remained unchanged through this assessment). Applying that utility function to the choice between alternative #1 and alternative #2 results in #2 being selected.

But #2 is highly risky, as can be seen from Figures 3.10 and 3.12 or from the risk premium of \$15.37M. The chances of #2 costing more than \$70M is 30%, much more than the 10% of alternative #1. Indeed, one could end up with costs of \$90M or even \$100M with #2, costs which are not possible with #1. This example shows that the utility of an alternative's PDF (if one accepts the utility function assessed from DOE technology managers) may not accurately capture all the potential risk in an operational, rather than theoretical, setting.

This can be illustrated by another example. If the cost PDF from alternative #1 is shifted down by \$20M, the resulting PDF is displayed in Figure 3.13. The shape of the cost distribution is the same, implying the same level of uncertainty in remediation costs.

The mean cost is \$45M, as one would expect, but the expected utility of #3 is 0.966. The associated CE is \$48.62M, yielding a risk premium of \$3.62M compared to \$2.05M for alternative #1. This would imply that the perceived risk increased, despite the fact that the costs are lower! While it is clear that alternative #3 would be preferred to #1 and #2, the way risk is indirectly measured in the utility function does not seem to clearly express our definition of risk.

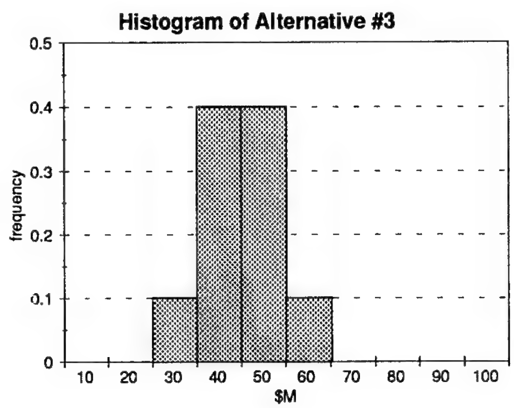


Figure 3.13

Further problems with risk expressed through utility result from the subjective nature of utility functions. A utility function represents the values of one person — the decision maker whose preferences were assessed through procedures like those mentioned in Chapter II. These preferences are captured at the time the utility function is assessed. While one can attempt to generalize the utility function to other times and different people, the only thing it unequivocally represents is the decision maker's preferences at the moment it was assessed.

For these reasons, utility functions alone are not the single best way to quantify and compare risk as one moves from the theoretical to the operational. Objective measures are needed that more directly measure what we define as technical risk.

3.4.3 *Mean and Range of an Attribute.* One way to condense the objective information contained in the histogram is to take the smallest, largest, and mean value displayed on it. This expresses the most likely or expected value of the represented PDF and shows the maximum variation about that expected value in both directions. While this is valuable information for the decision maker, information regarding the likelihood of the variations is left out. Values near the limits may occur with extremely low probability, thus misleading the decision maker as to the complete risk involved.

The *DPL*© software presents the results of an analysis through histograms of discrete cumulative probability distributions (CDFs) or probability distribution functions (PDFs). This presents some difficulty in examining a model's results, since the potential outcomes are represented in sets of intervals or bins. When simulation is used in DPL©, the actual outcomes of the different replications are not available — only the histograms are provided. Instead, each replication is approximated by the midpoint of its respective histogram bin [Mykytka, 1996b].

In such a setting, the lower and upper bounds of the attribute's range become midpoints of the lowest and highest bins from the histogram. This may under-represent the actual bounds by some small amount related to the number of bins used to form the histogram. Thus, the limits of the range of the PDF are only approximations of the true range of that attribute.

Calculations of the mean face similar difficulties. Let us say that n is the number of replications made for a given technology portfolio, and h is the number of histogram bins

or intervals chosen before running the DPL© model. Instead of summing up the replications and dividing by n, a different approach is required. If x is the attribute we are concerned about, the sample mean of this PDF of x is approximated by

$$\bar{x} \approx \sum_{i=1}^{h} xm_i \times p_i \tag{3.4}$$

where $\bar{x}$ is the sample mean, xm_i is the midpoint of the i^{th} histogram bin, and p_i is the relative frequency of occurrence of the i^{th} bin. This equation assumes that the width of the histogram bins is equal throughout the PDF of the attribute x.

The high, low, and mean values can be easily found using a spreadsheet with imported *DPL©* histogram files. Once the range and mean have been found for several alternative technology portfolios, they can be compared on a single graph far more easily than their parent histograms could be.

3.4.4 *Variance and Expected Unfavorable Deviation.* An alternative way to describe the PDF of the attribute of interest is through its variance about the sample mean. This also condenses information found in the histogram to a simpler form, but instead of representing the complete range of the attribute, the variance or its square root, the standard deviation, provides a sense of how the attribute is distributed without full knowledge of its range. Both consequence and probability are accounted for in a fashion.

While the sample variance is typically defined as

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1} \tag{3.5}$$

where the actual i^{th} replication is x_i [Mendenhall, et. al., 1990:343], we know that we cannot obtain the set of $\{x_i\}$ from *DPL*©. We therefore again adopt the midpoints of the histograms. The sample variance, based on the histogram midpoints, is then estimated by

$$S^2 \approx \sum_{i=1}^{h} (xm_i - \bar{x})^2 \times p_i. \tag{3.6}$$

If written in a form equivalent to Equation 3.5 when the set of $\{x_i\}$ is known, this formula uses a numerator of n instead of n - 1 [Mykytka, 1996a]. This is easy to see if one restricts the histogram bins to only one instance each. Then $h = n$ and $p_i = 1/n$. When we are using the simulation option of *DPL*© instead of full enumeration because of the size of the model involved, S^2 from Equation 3.6 is a biased estimator of the population variance (which would otherwise result from the actual full enumeration of the entire model). To correct for this, multiply the results of Equation 3.6 by $\frac{n}{n-1}$.

There is a potential problem when using variance or the standard deviation to represent risk, however. We are defining risk through the negative or unfavorable consequences and their likelihoods, and the variance counts deviations from the mean both in our favor and against. If the PDF is asymmetric, the variance may not be a good measure of technical cost and schedule risk. Instead, a measure of variation that counts

only the unfavorable departures from the mean should be used [Jia and Dyer, 1995:3; Weber, et. al., 1990].

Such a measure is the *expected unfavorable deviation*, or EUD.[1] It is similar to in concept to Jia and Dyer's "standard measure of risk" [1995:3], but is an objective measure rather than based on a utility function. It is defined as

$$\begin{aligned} EUD &= \sum_{i=1}^{n} \begin{cases} |x_i - \bar{x}| \times p_i & \text{when } x_i - \bar{x} \text{ is unfavorable} \\ 0 & \text{otherwise} \end{cases} \\ &\approx \sum_{i=1}^{h} \begin{cases} |xm_i - \bar{x}| \times p_i & \text{when } xm_i - \bar{x} \text{ is unfavorable} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{3.7}$$

This EUD is related to the semi-variance discussed in Chapter II, which is calculated in a similar way as the sample variance of Equation 3.6 but includes only the unfavorable variations. One can see that the semi-variance is almost the square of the EUD, but each term differs by a factor of p_i inside the summation.

Either will enable us to quantify the cost and schedule risks of the candidate portfolios by providing a numerical measure of the risk. The shape, not the location, of the attribute's PDF determines the EUD or semi-variance. By correcting for the PDF's expected value, the resulting statistics are independent of the mean of the attribute. This

[1]"Unfavorable deviation" rather than "negative deviation" is used here to avoid confusion. In some cases, such as cost and schedule, it is the deviations above the mean that are of concern (i.e. $x_i - \bar{x} > 0$) while in others, such as maximum speed or cargo capacity, it is the deviations below the mean (i.e. $x_i - \bar{x} < 0$).

allows one to use both the mean and the EUD or semi-variance to compactly represent the PDF of the attribute while preserving the information of most interest to decision makers.

The sample variance, semi-variance, and EUD can be calculated in a spreadsheet in much the same fashion as the sample mean is, using the histogram of the attribute's PDF. Equation 3.6 will result in S^2 using the histogram bin midpoints, while Equation 3.7 will generate the EUD. Note that the sample mean is required.

3.4.4.1 *EUD Example.* To illustrate the use of the expected unfavorable deviation to quantify risk, let us examine the past examples of section 3.4.2.1. For this illustration we will restrict ourselves to alternative #1, from Figure 3.9. The mean cost is \$65M, found using Equation 3.4. Since higher costs are undesired, the EUD is found to be \$3.5M using Equation 3.7:

$$
\begin{aligned}
EUD &= \sum_{i=1}^{4} |x_i - \bar{x}| \times p_i \qquad \text{when} \quad x_i - \bar{x} > 0 \\
&= 0 + 0 + (70 - 65) \times 0.4 + (80 - 65) \times 0.1 \\
&= 3.5 .
\end{aligned}
$$

In a similar fashion, the EUD of #2 (Figure 3.10) is \$7.25M and the EUD of #3 (Figure 3.13) is \$3.5M. Clearly #2 is riskier than either #1 or #3, while #1 and #3 have the same amount of cost risk. This agrees with the intuitive impression one gets from looking at the PDFs.

3.4.4.2 *EUD vs. Semi-variance.* It is hard to choose between semi-variance and EUD as measures of risk. In general, one may want to use semi-variance when one's expected audience or customer is knowledgeable about statistics and portfolio

analysis, and therefore used to seeing variances and standard deviations. When one's audience or customer is not familiar with the concept of variance, EUD is easier to explain, being a linear function of $|x_i - \bar{x}|$, and in the same units as the attribute of interest.

Semi-variance and EUD will not necessarily produce the same results, however, given the same data. While one might expect the two risk measures to be functionally equivalent, ranking the same set of alternative in the same order, this may not occur. This can be demonstrated by an example.

Let us examine two different alternatives, represented by discrete PDFs where there are only two points above the mean for each (assuming that above the mean is undesirable). In these cases, the EUD and semi-variance for the j^{th} alternative are:

$$\begin{aligned} \text{EUD}_j &= (x_{1j} - \bar{x}_j)\cdot p_{1j} + (x_{2j} - \bar{x}_j)\cdot p_{2j} \\ \text{SV}_j &= (x_{1j} - \bar{x}_j)^2\cdot p_{1j} + (x_{2j} - \bar{x}_j)^2\cdot p_{2j} \end{aligned} \tag{3.8}$$

where x_{ij} represents the i^{th} point above the mean for the j^{th} alternative, p_{ij} is the probability of getting x_{ij}, and $\bar{x}_j \le x_{1j} < x_{2j}$. The possibility of generating different risk rankings could only occur if $\text{EUD}_1 > \text{EUD}_2$ when $\text{SV}_1 < \text{SV}_2$ (or vice versa). Since $\bar{x}_j$ is a constant, let $a_i = x_{i1} - \bar{x}_1$ and $b_i = x_{i2} - \bar{x}_2$. Then, looking at the case where $\text{EUD}_1 > \text{EUD}_2$ and $\text{SV}_1 < \text{SV}_2$, the possibility of different risk rankings can only occur if:

$$\begin{aligned} a_1\cdot p_{11} + a_2\cdot p_{12} &> b_1\cdot p_{21} + b_2\cdot p_{22} \\ a_1^2\cdot p_{11} + a_2^2\cdot p_{12} &< b_1^2\cdot p_{21} + b_2^2\cdot p_{22} \end{aligned} \tag{3.9}$$

For this example, let $p_{11} = p_{21}$ and $p_{12} = p_{22}$. Then Equation 3.9 becomes:

$$\begin{aligned} ka_1 + a_2 &> kb_1 + b_2 \\ ka_1^2 + a_2^2 &< kb_1^2 + b_2^2 \end{aligned} \tag{3.10}$$

where $k = \dfrac{p_{11}}{p_{12}} = \dfrac{p_{21}}{p_{22}}$. Focusing our attention on b_2, Equation 3.10 becomes

$$\begin{aligned} b_2 &< ka_1 + a_2 - kb_1 \\ b_2^2 &> ka_1^2 + a_2^2 - kb_1^2 \end{aligned} \tag{3.11}$$

Since $b_2 > 0$ and assuming $ka_1^2 + a_2^2 > kb_1^2$,

$$\begin{aligned} ka_1 + a_2 - kb_1 < b_2 &< \sqrt{ka_1^2 + a_2^2 - kb_1^2} \\ \therefore\ ka_1 + a_2 - kb_1 &< \sqrt{ka_1^2 + a_2^2 - kb_1^2} \end{aligned} \tag{3.12}$$

Equation 3.12 implies that ranking differences for this case can occur if a_1 and/or b_1 is sufficiently less than 1.

The condition represented by Equation 3.12 is possible — Figures 3.14 and 3.15 show a comparison between two two-point alternatives where x_{22} is allowed to change. Here it varies between 0.8 and 0.82. As x_{22} increases, EUD_2 and SV_2 also increase. Since x_{11} and x_{12} are constant, the first alternative's EUD and semi-variance are constant at 0.5 and 0.388, respectively. The intersections of the two EUD and semi-variance lines differ, showing a region of between about 0.803 and 0.817 where EUD1 > EUD2 but SV1 < SV2.

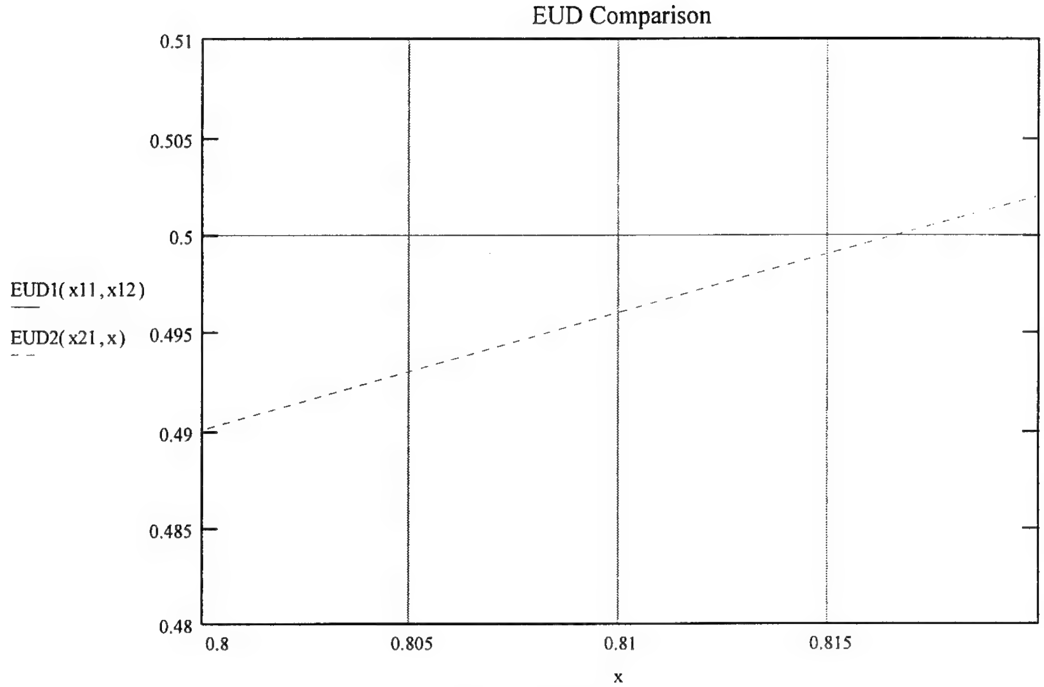


Figure 3.14

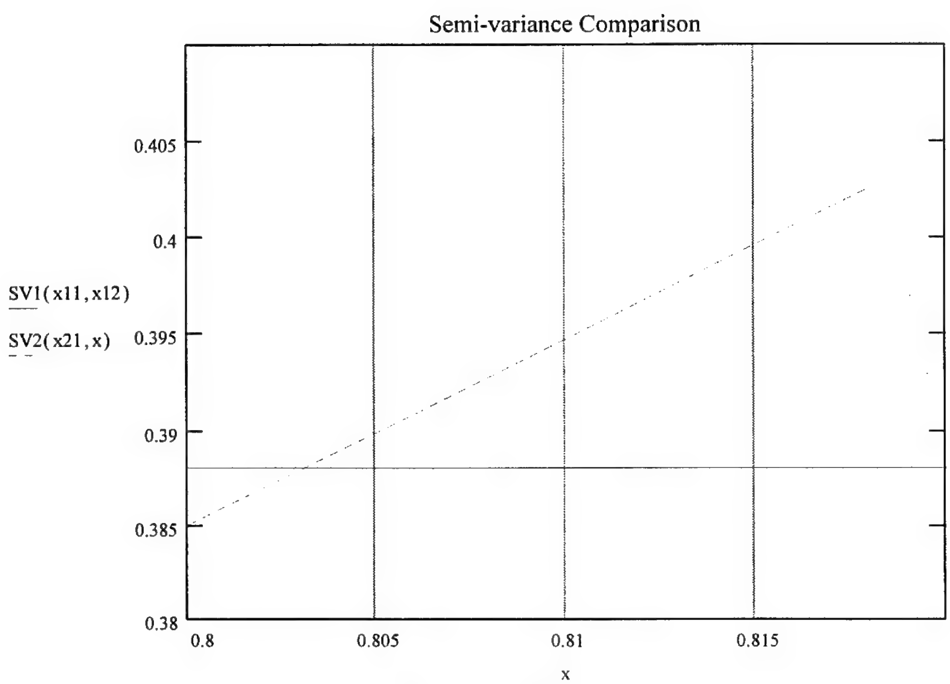


Figure 3.15

This potential for ranking differences has its cause in the squaring of the deviation in the semi-variance formula. When $x_i - \bar{x}$ for the i^{th} occurrence is less than one, the contribution to EUD is more than that to the semi-variance. This is the opposite of what happens when $x_i - \bar{x}$ is greater than one. This is a complication of some concern and is further motivation to use the EUD rather than the semi-variance as a measurement of risk. EUD remains a consistent measure across the range of $x_i - \bar{x}$, while the semi-variance may behave differently dependent on what units are used.

3.4.5 *Summary of Histogram Measures.* To review the risk measures developed from the output histograms, consider Figure 3.16 and Table 3.3. This cost histogram is typical of the pilot study results, being highly asymmetric with some small frequency of

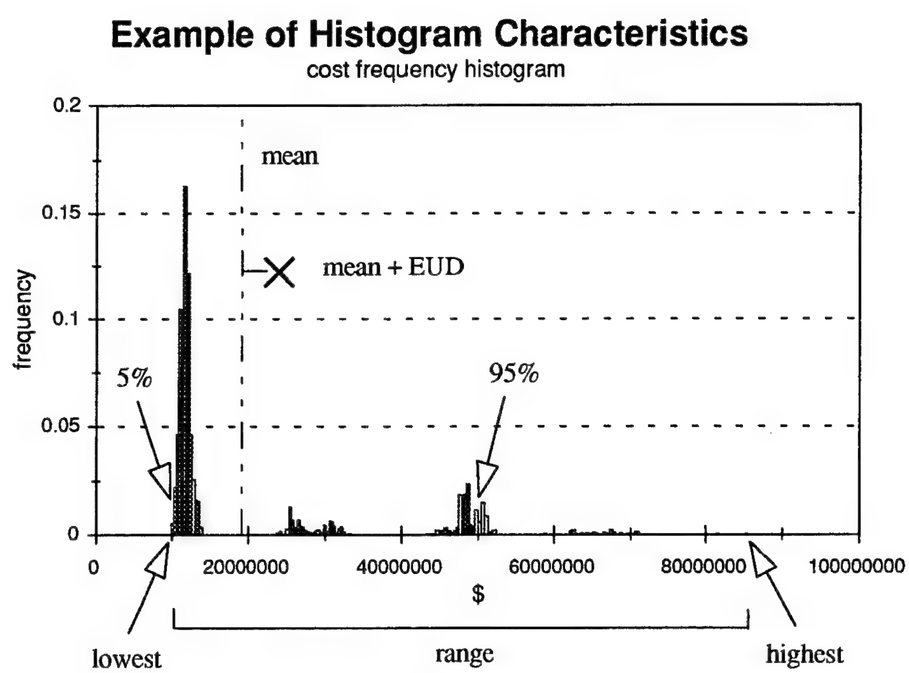


Figure 3.16

extraordinarily high results. The term (mean + EUD) is shown for later reference with the Chapter IV results. The variance and semi-variance are not displayed to preserve clarity. Note how the 95% fractile point is far from the actual highest cost.

Summary of Histogram Features

feature	what it measures
mean	expected value of PDF
range	spread of PDF
low	spread below the mean
high	spread above the mean
5% fractile	spread below the mean
95% fractile	spread above the mean
variance	general deviation from mean
semi-variance	downside risk
EUD	downside risk

Table 3.3

3.5 *Summary of Methodology*

A review of the alternatives and decisions of the methodology described in Chapter II shows how concepts from the literature and careful analysis of the DOE's remediation technology problem are used in the decision support system. The combination of risk assessment and technology forecasting can be broken down into dealing with model inputs or outputs.

3.5.1 *Model Inputs*. Cost and schedule risks involved with research and development efforts are modeled by soliciting expert opinion for subjective probability

Risk: Time to Complete Development
Method: Release Date Distribution

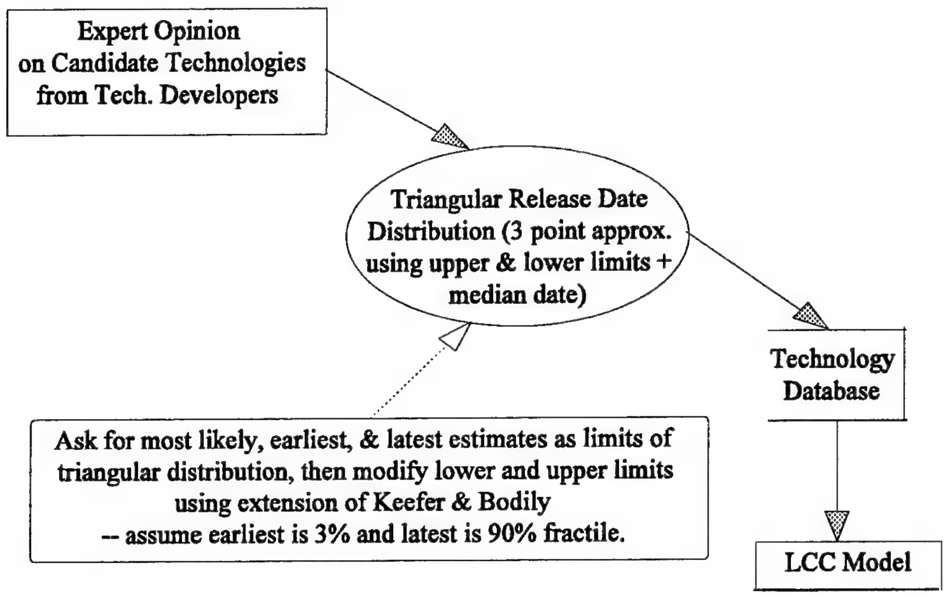


Figure 3.17

distributions of the dates the technologies are released from R&D. These release date distributions take the form of triangular distributions, using three parameters of earliest, most likely, and latest possible time from the present to be fully specified. Because of concerns about under-representing the extremes of these distribution, the tails are extended by assuming the expert's estimates of the earliest and latest dates are actually the 3% and 90% fractiles and adjusting the distributions accordingly. The total R&D costs are then estimated by multiplying this release date distribution by a constant annual cost drawn from current project projections (see Figures 3.17 and 3.18 for process action

Risk: Cost to Complete Development
Method: Cost as a Function of Release Date

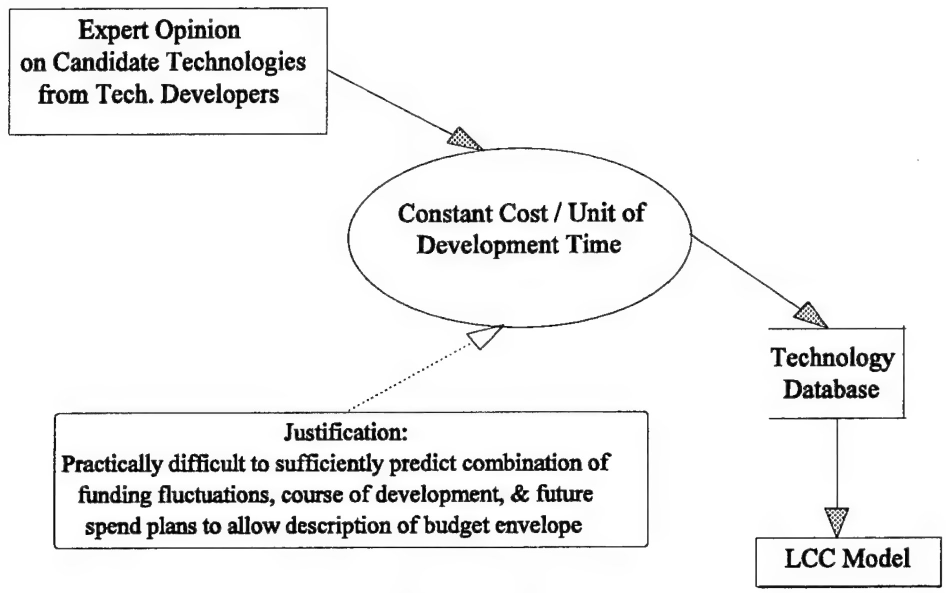


Figure 3.18

diagrams[2] graphically depicting what is being done).

The performance of technologies in the field is represented by random variables drawn from expert opinion and used in the LCC Module. The possibility of the technology completely failing in the field is accounted for by expert judgement of the probability that the technology fails to perform as expected, given that the preliminary landfill characterization may not necessarily correct and that the technology successfully completed R&D (see Figure 3.19).

[2]The open box "Technology Database" refers to the data store used to hold technology information (see Figure 3.1) using the process action diagram notation in Shina, 1991 [14-16].

Risk: Chance that Tech. Fails In the Field
Method: Expert Estimates of P(use) per Technology

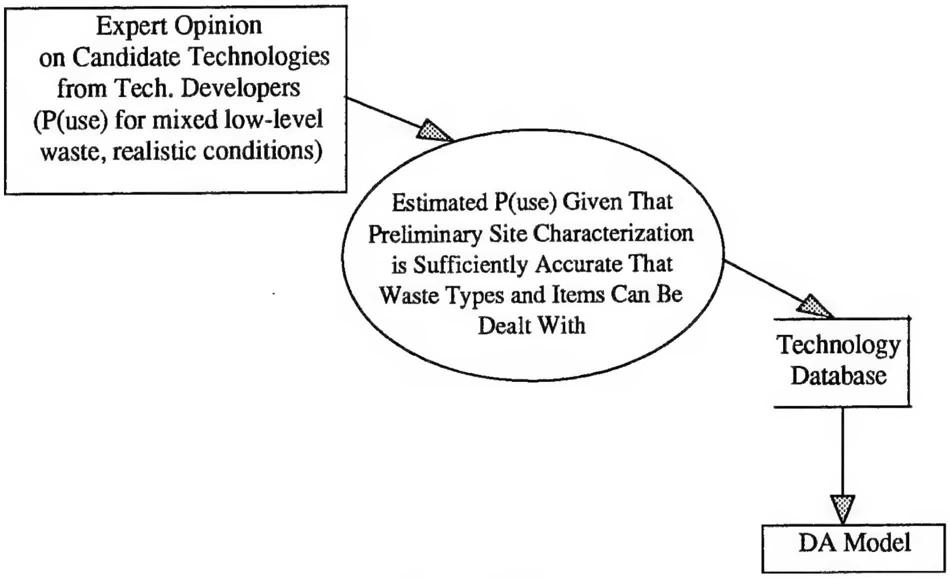


Figure 3.19

The performance of technologies in the field is represented by random variables drawn from expert opinion and used in the LCC Module. The possibility of the technology completely failing in the field is accounted for by expert judgement of the probability that the technology fails to perform as expected, given that the preliminary landfill characterization is not necessarily correct and that the technology successfully completed R&D (see Figure 3.19).

The risk that a given technology cannot meet regulatory requirements governing the remediation of that specific waste site is too complex and site specific to be modeled in the decision support system. Instead the user of the model is asked to make this judgement based on his or her greater understanding of the specific site being examined.

3.5.2 *Model Outputs*. The technologies are employed in complete portfolios to conduct the entire remediation of a landfill in the Decision Analysis Module, using information about the R&D and operational schedule and costs drawn from expert opinion and the LCC Module. The DA model creates output distributions of total cost and time for each portfolio using simulation, and recommends the best portfolios based on a multi-attribute utility function for cost and schedule.

These resulting distributions can be examined to find expressions of the risks of these alternatives. The range and mean provide one way to present the information contained in the output probability distributions. While the utility scores of the alternatives implicitly include risk, a more operational measure of risk is desired. This is provided by the semi-variance or expected unfavorable deviation (EUD), which numerically expresses the risks of cost and schedule overruns so that portfolios can be quantitatively compared.

IV. *Results*

This chapter will describe the results of applying some of the concepts and methods previously developed. The prototype Decision Analysis Module was used with incomplete technology information gathered from the technology developers, supplemented with notional data, to demonstrate its features and test the concept. The input data and the resulting portfolio schedule and time distributions were examined using the procedures from Chapter III. This provides examples to guide later use of the overall decision support model and demonstrates ways to see the cost, schedule, and performance risks of recommended technology decisions.

4.1 *Preliminary Technology Information*

A complete prototype for the overall decision support system is scheduled for completion by the summer of 1996. Information is being gathered by MSE on two to three different technologies for each remediation process to demonstrate the prototype to DOE/EM-55 in October 1996. Interviews with the principle investigators of each technology development project by MSE personnel were originally planned for the fall of 1995, however faxed questionnaires were used instead (the interview script is attached in Appendix D). The gathering of this information, a responsibility of MSE, has not been completed at this point (March 96). However, some initial survey results supplemented with the expert opinion of MSE personnel were used to pilot test the Technical Risk and the Decision Analysis Modules. The data should be treated as notional and used for proof

of concept only. The preliminary technology data relevant to the Technical Risk Module is attached (see Appendix A).

4.1.1 *Adjusting R&D Release Date Distributions.* The preliminary release dates were solicited from the principle investigators and MSE by requesting estimates of the earliest, most likely, and latest possible dates, measured in years from the present. As described in Chapter II, these release dates are expected to be conservative, resulting in a triangular distribution that has unrealistically small tails. The procedure described in Chapter III was used to adjust the range of the distributions to include more of the low

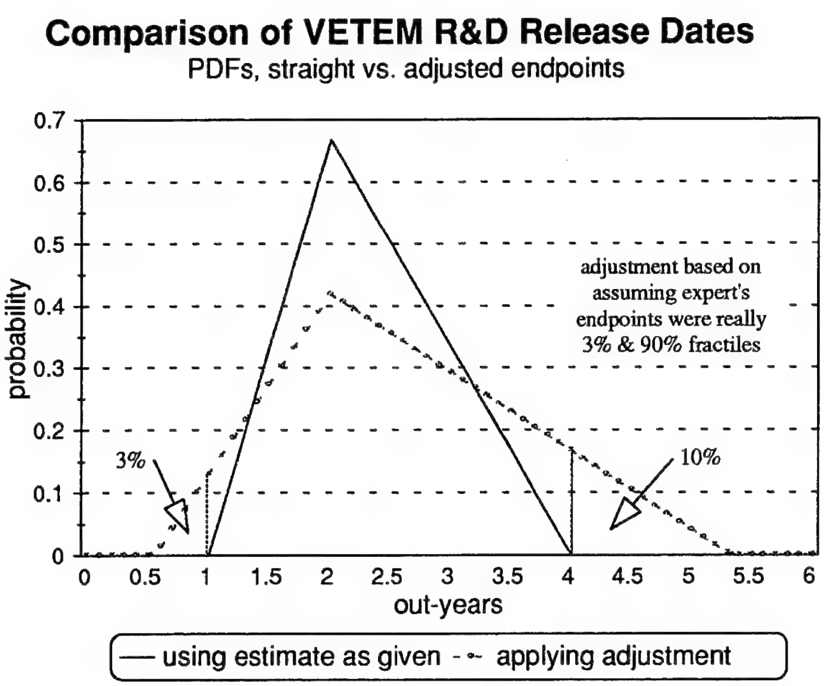


Figure 4.1

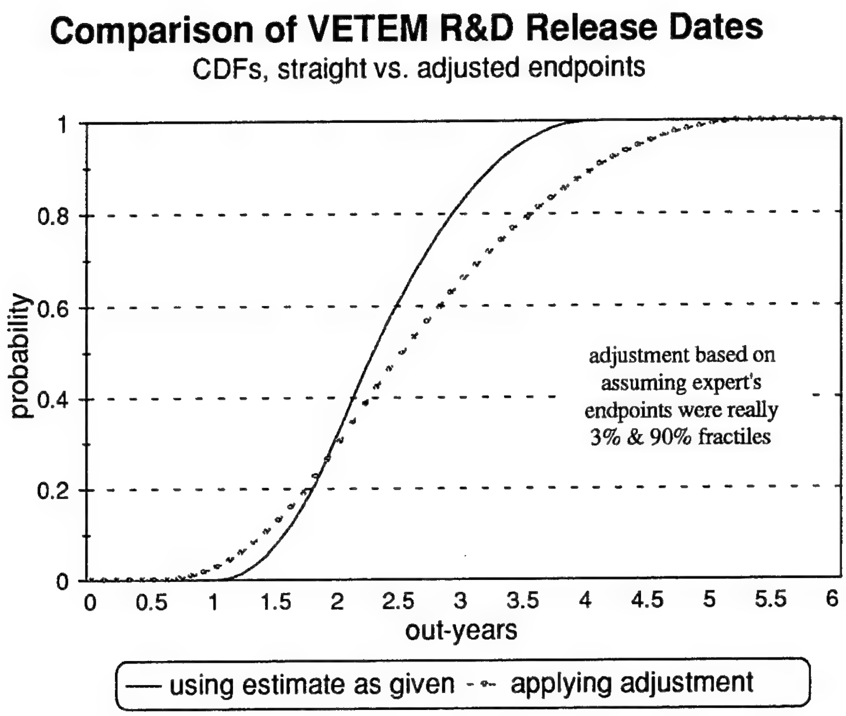


Figure 4.2

probability possibilities. A simple *MathCad*© 5.0+ file was used to solve the simultaneous equations, with the "SmartMath" option enabled (attached in Appendix E). This results of this procedure are shown in Figures 4.1 and 4.2 for the second characterization technology, VETEM. The adjusted release date limits for all technologies are included in Appendix B.

The greatest increase is, of course, in the latter part of the distributions, since we are assuming that the expert-provided latest date is actually the 90% fractile (recall that the expert's earliest date estimate is assumed to be the 3% fractile). The feasible solution to Equation 3.1 moves the earliest and latest dates from 1 and 4 years to 0.549 and 5.330 years, respectively. The total range of the release date distribution increases from 3 to

4.781 years, an increase of almost 60%. While this may seem like a large increase, because the likelihood of these dates occurring is small, the mean date changed very little — going from 2.333 to only 2.626 years. The variance, however, increased from 0.389 to 1.001, due to the spreading of the distribution.

Similar results were found when adjusting the other release date distributions in the preliminary technology database. Means increased by an average of only 9% after this procedure was used, while the variance increased by an average of 141%. These increases in variance underscore the need for accurate estimates.

4.1.2 *Estimates of Annual R&D Costs*. Based on the preliminary information gathered or generated by MSE, the total remaining development costs for the set of technologies being examined were estimated and are given in Appendix A. These figures, divided by the mean from the adjusted release date distribution, provide an estimate of the annual R&D cost for that development project. This will be used in the LCC simulations to determine the simulated R&D cost for a given draw from the release date distribution and are also listed in Appendix B.

The annual R&D cost estimates are lower when using the adjusted release date distributions instead of the release dates of MSE, because the mean release dates increased. The total R&D costs remain the same as shown in Appendix A.

4.1.3 *Estimates of the Probability of Successful Field Use*. The probability of successful use in the field, P(use), was estimated by MSE for all the technologies included in the future prototype demonstration. Since the landfill being considered holds mixed

low-level waste [Nickelson, 1996], P(use) was defined as the probability that the technology would work as expected at a mixed waste landfill given the normal uncertainty in preliminary characterization and assessment of the site.

The accuracy of these point estimates is uncertain. Without actual performance data or information on the past accuracies of preliminary assessment efforts, anything other than subjective opinion about the future performance of these technologies is difficult to find. The sensitivity of portfolio selection to changes in P(use) will be examined in this pilot study and is strongly recommended for any future use of the overall decision support system. These estimates, while notional, are adequate for this demonstration.

4.2 *Examination of Preliminary Results*

Because the LCC Module is not yet complete, simulations of the operating cost and schedule distributions were not available. To allow the exercise of the Decision Analysis Module, MSE personnel provided assessments of the cost and schedule distributions for each candidate technology. Appendix A shows these notional estimates. Ralston [1996] provides a complete description of this module.

A landfill at INEL in Idaho Falls, ID, was selected as the landfill requiring remediation. This landfill, Pit 9, was operated as a waste disposal pit from November 1967 to June 1969. One acre (43560 sq. ft.) was excavated to the basalt bedrock before being filled with approximately 150,000 cubic feet of packaged waste and 350,000 cu. ft.

of soil, then covered by 250,000 cu. ft. of overburden. This leaves 500,000 cu. ft. of mixed low-level waste to be remediated [Nickelson, 1996].

The DA model was run for two cases: 1) stabilization technologies were used in the remediation effort and 2) with the second characterization and the second monitoring technologies selected *a priori*, with stabilization excluded as an option. Because the decision to use stabilization is based on the results of the characterization and assessment process and judgement of the waste's stability and migration potential, we did not include the stabilization decision directly in the DA model. Instead, both stabilized and unstabilized strategies should be examined. For the unstabilized case, VETEM was arbitrarily picked as the characterization technology used from which the decision not to stabilize was made. The use of on-site monitoring was chosen because its cost and schedule distributions clearly dominated the Yucca Mt. option for the notional data employed in this study.

Two different pairs of cost and schedule utility functions are then required, one for the stabilized strategy and one for the non-stabilized strategy. These utility functions are shown in Figures 4.3-4.6. The two utility functions are combined via additive multi-attribute utility functions of the form:

$$u_{total}(cost,\ time) = k\ u_{cost}(cost) + (1 - k)\ u_{time}(time). \tag{4.1}$$

where $k = .667$ in both cases. These utility functions were assessed from interviews with technology managers working at the Landfill Focus Area Field Office at the Savannah River Site in South Carolina. They reflect the simple, but operational concept that the

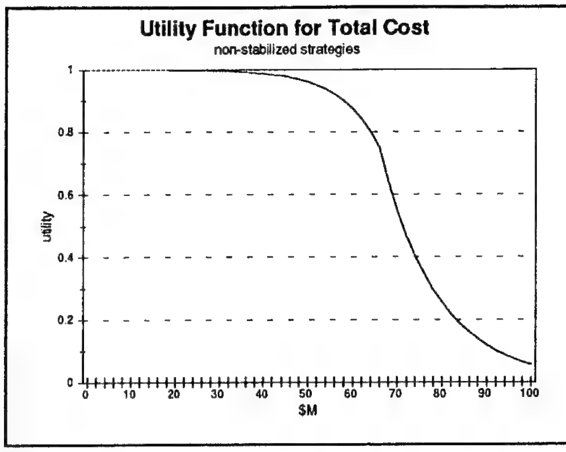


Figure 4.3

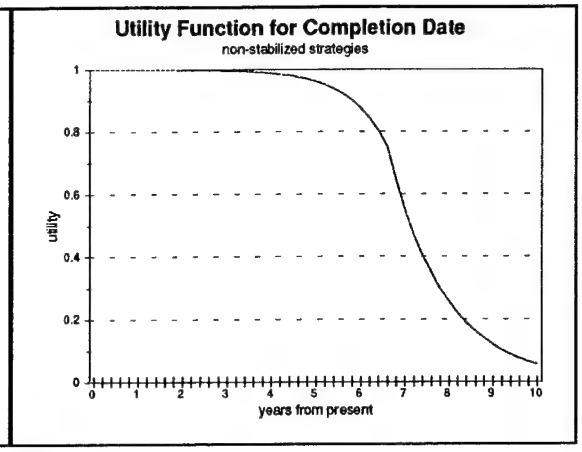


Figure 4.4

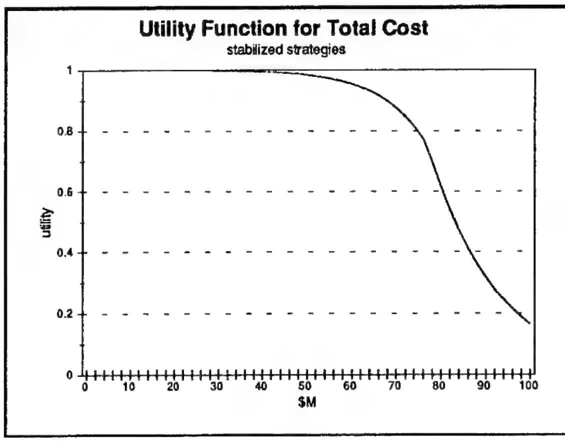


Figure 4.5

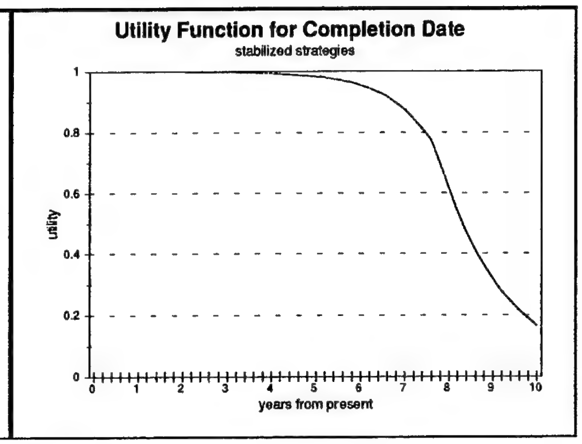


Figure 4.6

soonest completion date is preferred (see Appendix F for the actual equations).

After the stabilization decision is made, the decision paths break down into the ones shown on Figures 4.7 and 4.8. The upper paths correspond to cases where stabilization is used. The decision to pursue a containment vs. retrieval-treatment-disposal strategy is left open. Likewise, the bottom paths reflect the choice to not stabilize. A technology must be selected for each process in the chosen path. Because of the size of

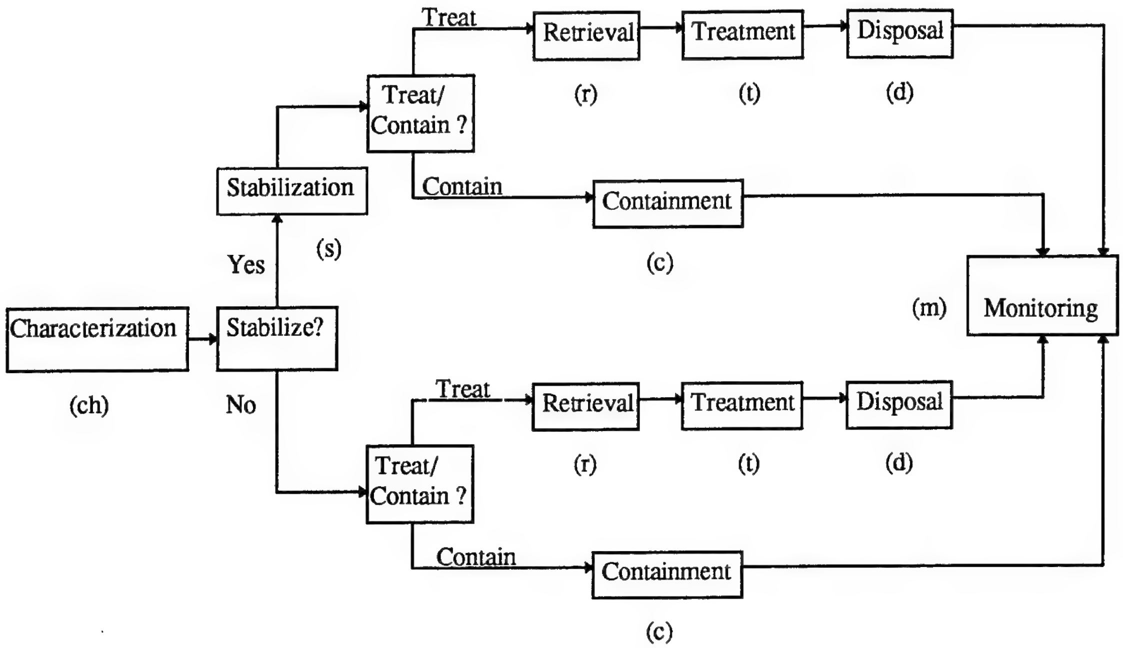


Figure 4.7

the model, it appeared prohibitive to completely enumerate all possible combinations of nodes in the DA model's decision tree. *DPL*©'s simulation option was used therefore with ten thousand iterations in each run instead of complete enumeration. Ten thousand iterations were felt to be sufficient to get accurate sample statistics.

The preliminary results found the best five strategies (as determined through total utility) for the two above cases. The technologies for these portfolios, one for each process, are listed in Table 4.1 using the ID codes found in Figure 4.7.

The processes in Figure 4.7 are not employed in a strictly sequential fashion. Some processes, specifically treatment, disposal, and monitoring, can begin while their predecessors are still underway if allowed by their R&D release dates. While in general

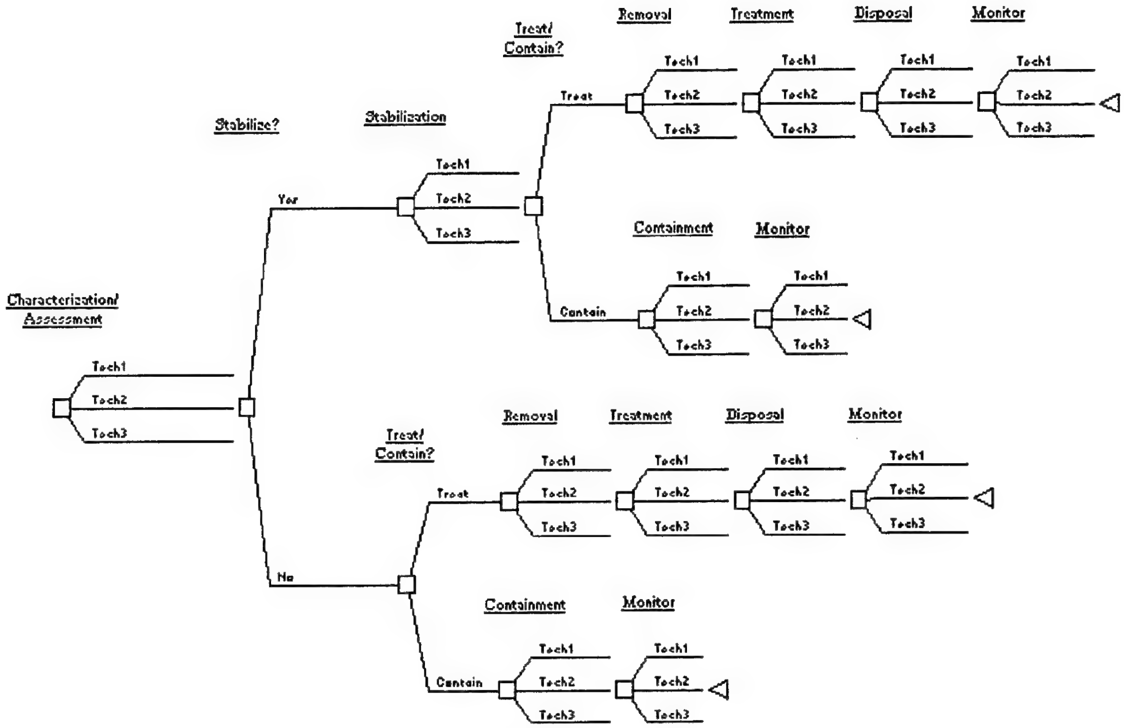


Figure 4.8

each technology is employed independently in the DA model, interactions between certain technologies from different processes are modeled, where one cannot be used with another or two technologies must be used together. Ralston discusses these factors in more detail [1996].

4.2.1 *Cost, Time, and Utility Histograms.* Examination of the total cost, schedule, and utility histograms resulting from the *DPL*© runs demonstrates the various risk measures described in Chapter III. Figure 4.9 shows a typical cost distribution, that of the #3 portfolio without stabilization from Table 4.1, while Figure 4.11 shows its time distribution and Figure 4.13 shows its utility distribution.

Both undesired consequences (higher costs, longer completion schedules, and lower utilities) and the probabilities of these events occurring are captured on these charts. Another way to view this information is through the cumulative distribution

Best Technology Portfolios Recommended By DA Module

When Stabilization Is Not Used

#1	ch2, cont1, m2
#2	ch2, r1, t1, d2, m2
#3	ch2, r2, t1, d2, m2
#4	ch2, cont3, m2
#5	ch2, r1, t3, d2, m2

When Stabilization Is Used

#1	ch1, s1, c1, m2
#2	ch2, s1, c1, m2
#3	ch3, s1, c1, m1
#4	ch2, s1, c3, m2
#5	ch3, s1, c3, m2

Table 4.1

functions, where the frequencies of occurrences are added together instead of plotted separately. This makes finding points such as the 5% and 95% limits easier. Figures 4.10, 4.12, and 4.14 show the cumulative distributions for the cost, schedule, and utility distributions in Figures 4.9, 4.11, and 4.13, respectively.

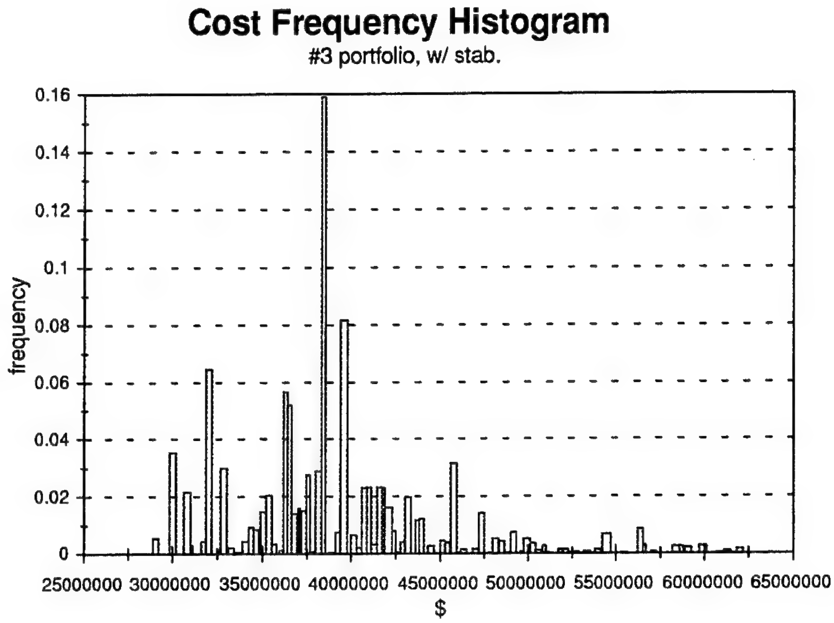


Figure 4.9

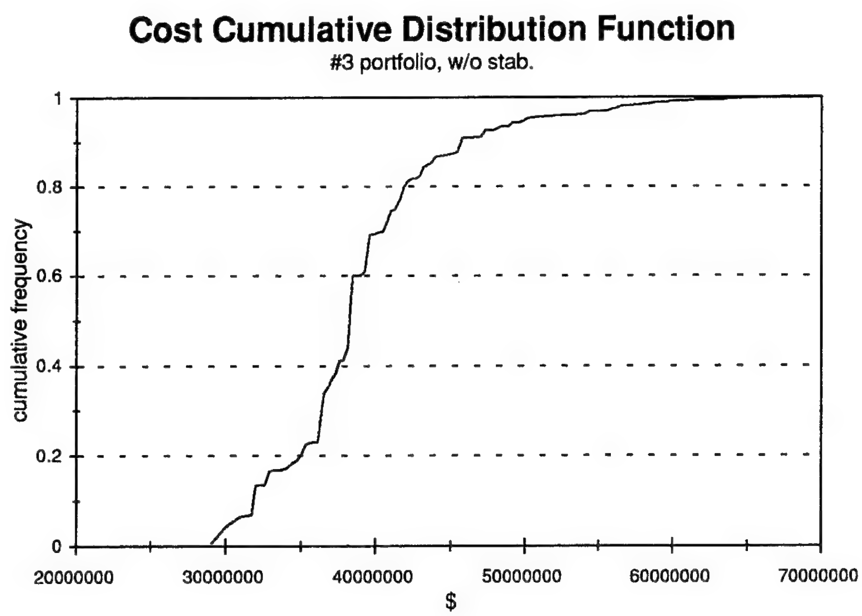


Figure 4.10

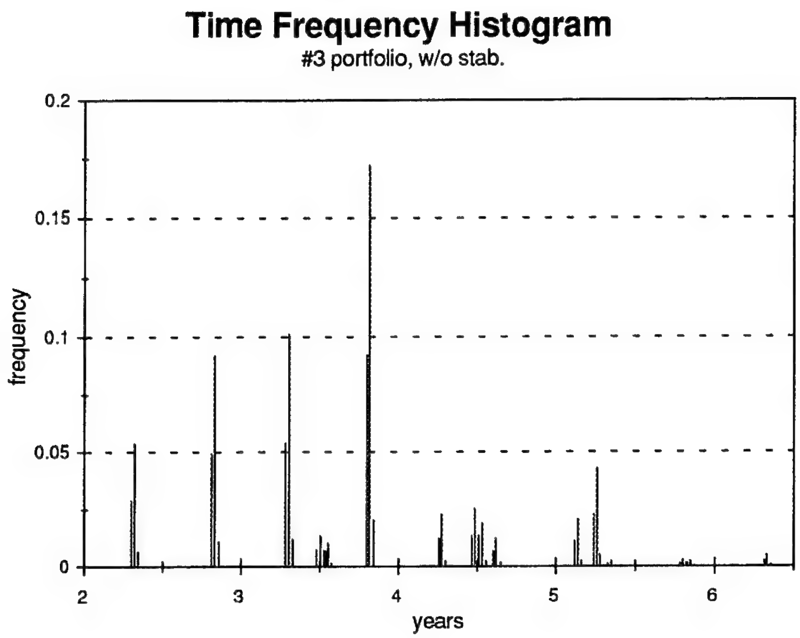


Figure 4.11

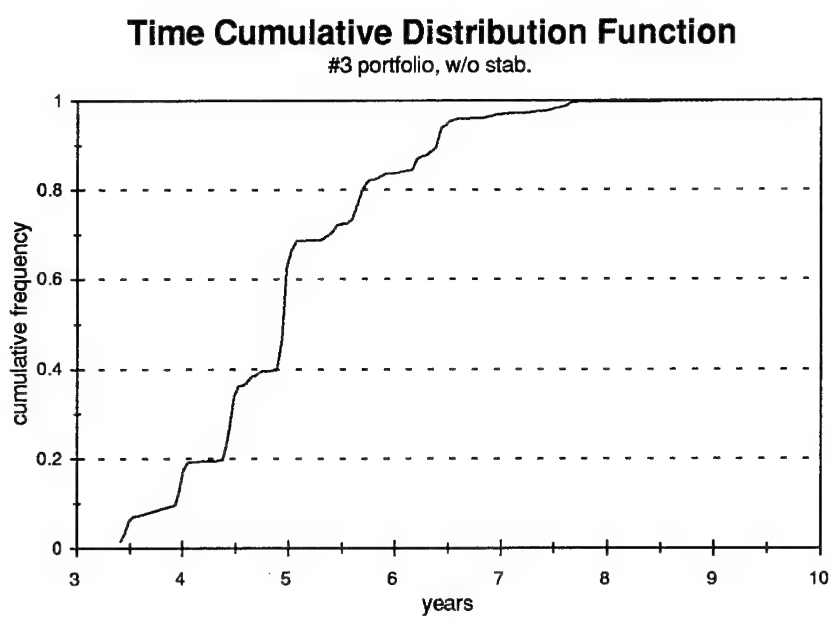


Figure 4.12

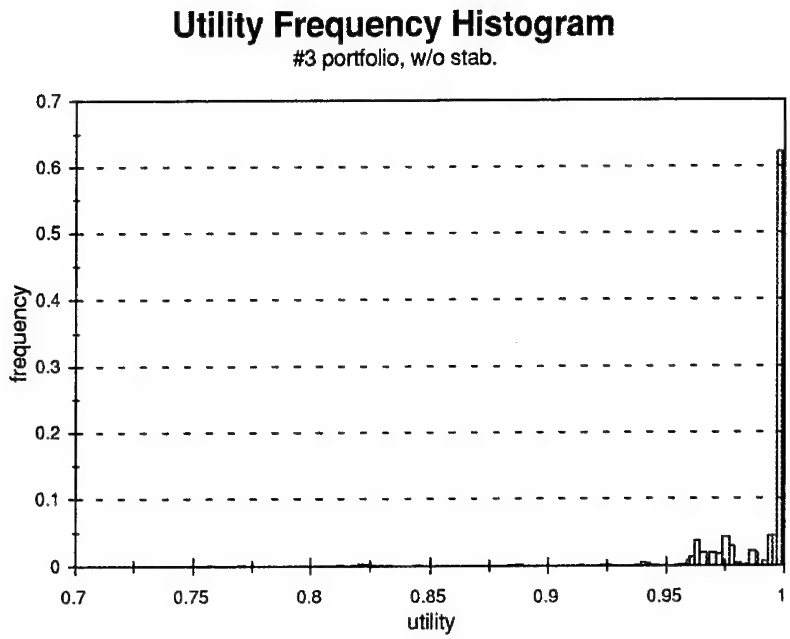


Figure 4.13

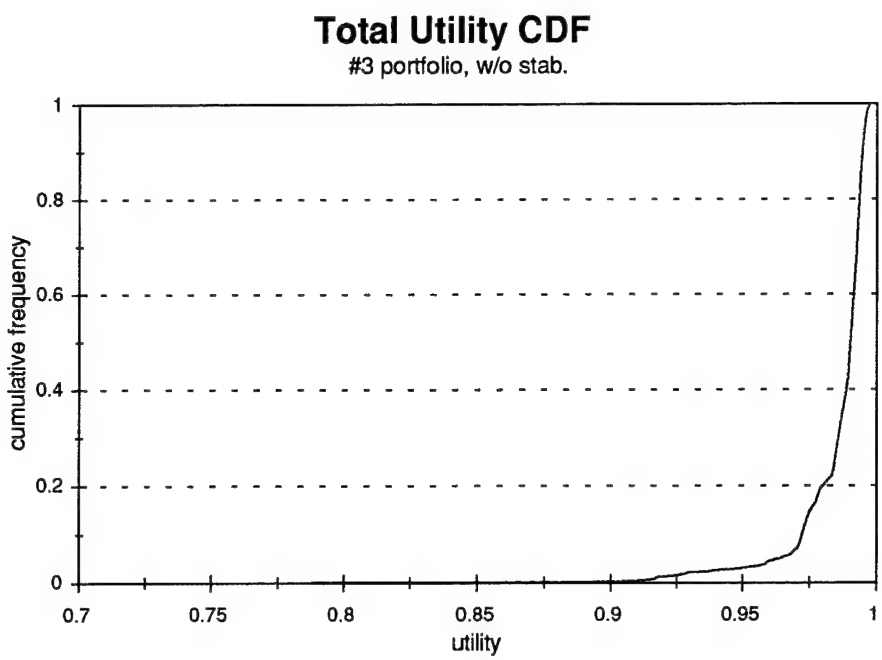


Figure 4.14

4.2.1.1 *DPL© Histogram Bins*. A careful review of Figure 4.9 will disclose an anomaly with this *DPL©* output. The widths of the histogram bars do not remain the same throughout the graph. This seems to be true for every result from the DA model that has bars of some width. While the reasons for this irregularity are unknown at this time (March 96), with the large sample size used in this study it does not seem to have a great effect on the results. See Appendix G for a discussion of this irregularity.

4.2.2 *Range Graphs*. Using the sample mean formula in Equation 3.4 and the largest and smallest histogram midpoints from the *DPL©* runs for the top portfolios listed in Table 4.1, we can plot the ranges of cost, time, and total utility for the cases with and without stabilization. From these plots we can understand the relative ranking of the technologies with respect to average cost and completion time and also see a measure of the risk of each portfolio. Figures 4.15-4.20 show these plots for the preliminary results.

As one can see from Figure 4.15, there is a dramatic difference in terms of range between the portfolios following removal-treatment-disposal strategies (#2, #3, and #5) and those that use containment (#1 and #4). From Figure 4.16, we can tell that the ranges of required time for completion are roughly the same for all five portfolios and that the means are what distinguish between them. Finally, the plot of utilities in Figure 4.17 shows the surprising low of zero utility for portfolios #4 and #5. This means that in at least one instance, the simulation of these portfolios resulted in breaking one of the cost or schedule constraints of the DA model and therefore being assigned zero value. A

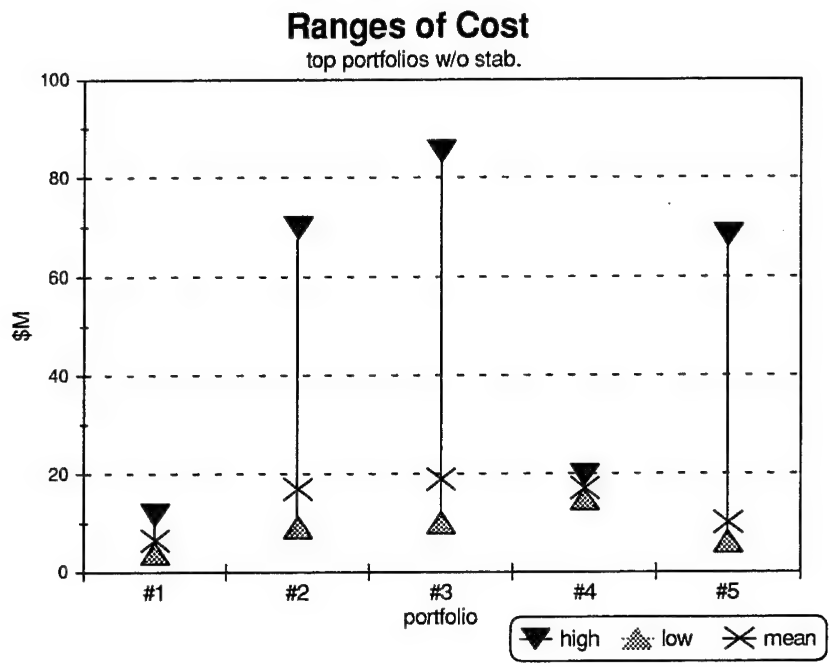


Figure 4.15

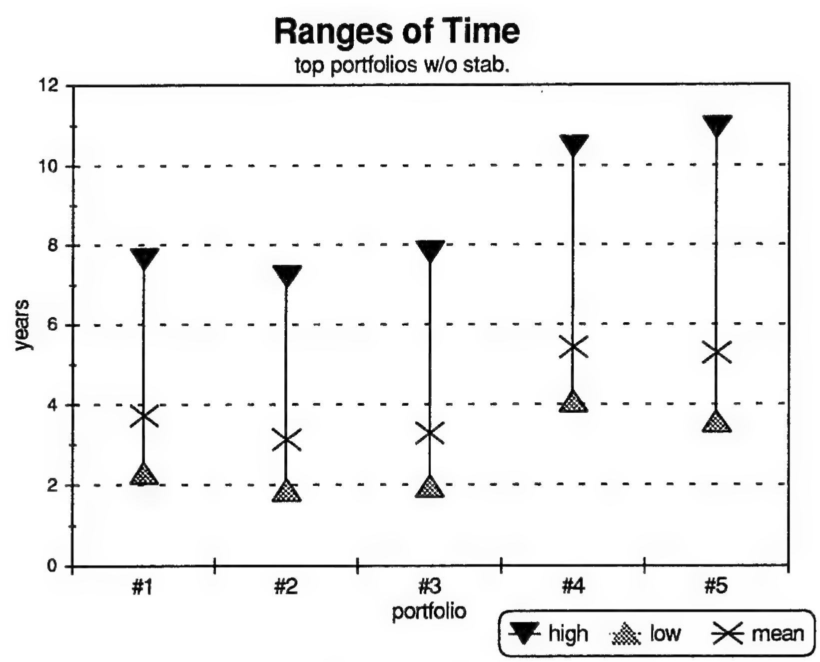


Figure 4.16

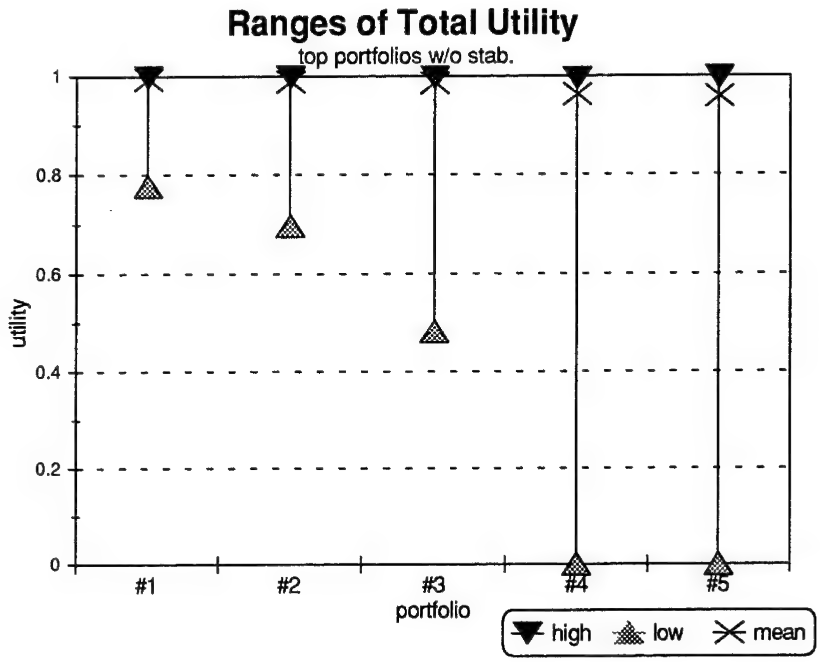


Figure 4.17

review of Figure 4.16 indicates that it was the schedule constraint of 10 years.

The portfolios following a stabilize-first strategy show fairly consistent ranges of cost, although the mean costs vary from \$40M to \$50M. A cursory examination of Figure 4.18 should cause one to wonder why portfolio #1 was ranked first by the DA model. Figure 4.19 identifies the reason — portfolio #1 has a dramatically shorter expected schedule. Since the ranges overlap, we know that there is no deterministic dominance involved. We would have to compare the original CDFs to determine the existence of stochastic dominance. This illustrates the trade-offs between the importance of cost and schedule implied by the constant k in the additive utility function of Equation 4.1 (page 4-6). We can also see the upper limit of completion time for #4 and #5 violates

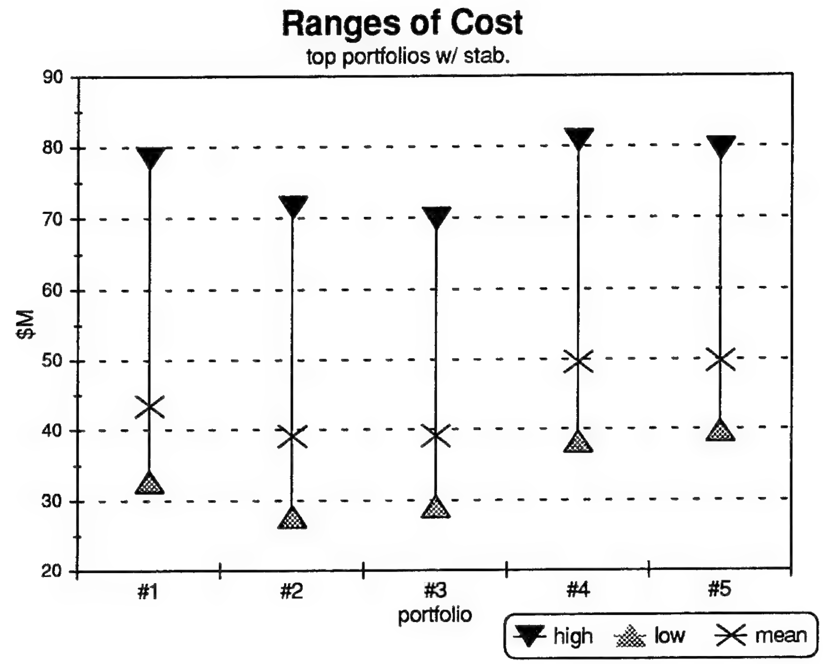


Figure 4.18

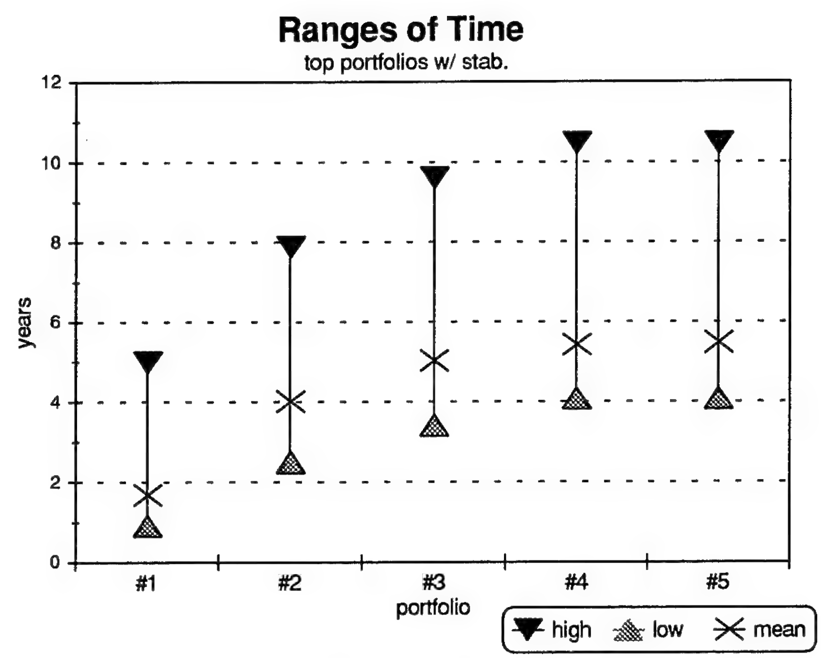


Figure 4.19

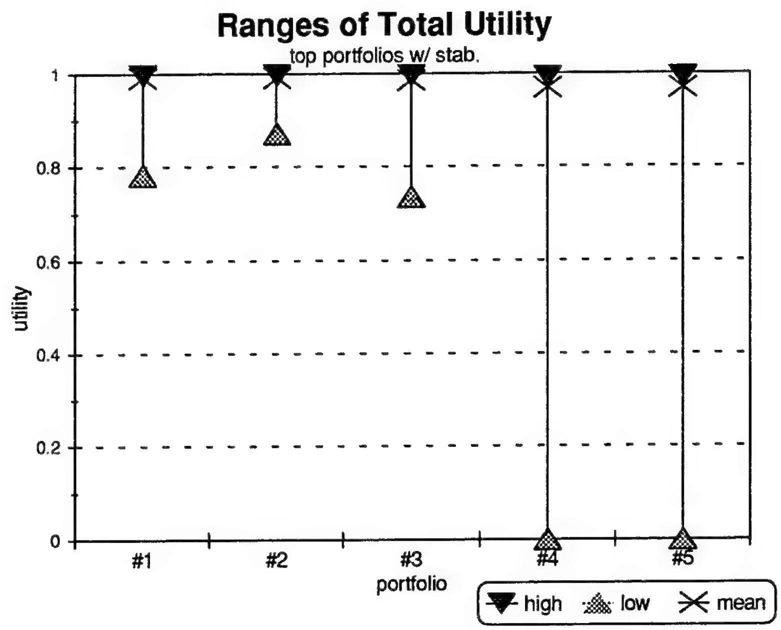


Figure 4.20

the 10 year constraint, resulting in a lower utility of zero for this set of runs as with the non-stabilized portfolios. Figure 4.20 could also make one wonder why portfolio #1 was ranked before #2, since #2's range of total utility is tighter than #1's. A check of the data in Appendix C shows that the difference in mean utilities is less than 0.0005 (#1: 0.99184, #2: 0.99180), indicating and highlighting that the tradeoff between cost and time for these portfolios is very close. Other factors, such as risk or political considerations, may then come into play to distinguish between the portfolios.

4.2.3 *Expected Unfavorable Deviations.* Similar graphs can be developed using the sample means and EUDs. While these do not represent the complete ranges of the cost and schedule results, they are a better representation of risk since probability is

incorporated in the definition of EUD (Equation 3.7). Figures 4.21-4.36 show the EUD graphs for the top portfolios. The actual numerical results are shown on Table 4.2.

Looking for risk with respect to utility may not be as meaningful to a decision maker as reviewing risks in tangible attributes of cost and schedule. Using the variance or EUD of a utility distribution also mixes two different types of risk definitions, that of classic utility theory and the "mean-variance" definition. Since the shape of the utility function determines, in part, the distribution of utility around the expected value for a portfolio, taking a measure of the variation around the mean "counts" the variation twice. Despite these theoretical cautions, however, this information is valuable to a decision maker trying to weigh the risks in a practical situation.

Figure 4.21 shows that the EUD measure is consistent with the ranges of cost for the non-stabilized portfolios. Portfolios #1 and #4 have very little expected variation from the mean values of $6.56M and $18.94M, respectively, while the retrieval-treatment-disposal portfolios (#2, #3, and #5) exhibit a great deal more cost risk. From Figure 4.22 we can see that all five portfolios have roughly equivalent schedule risks. The large cost EUDs imply that there is a great deal of uncertainty or variability in the preliminary cost estimates of retrieval, treatment, and disposal technologies. The utility means on Figure 4.23 decrease going from #1 to #5 (since that is what was used to rank order the portfolios), and the EUDs increase.

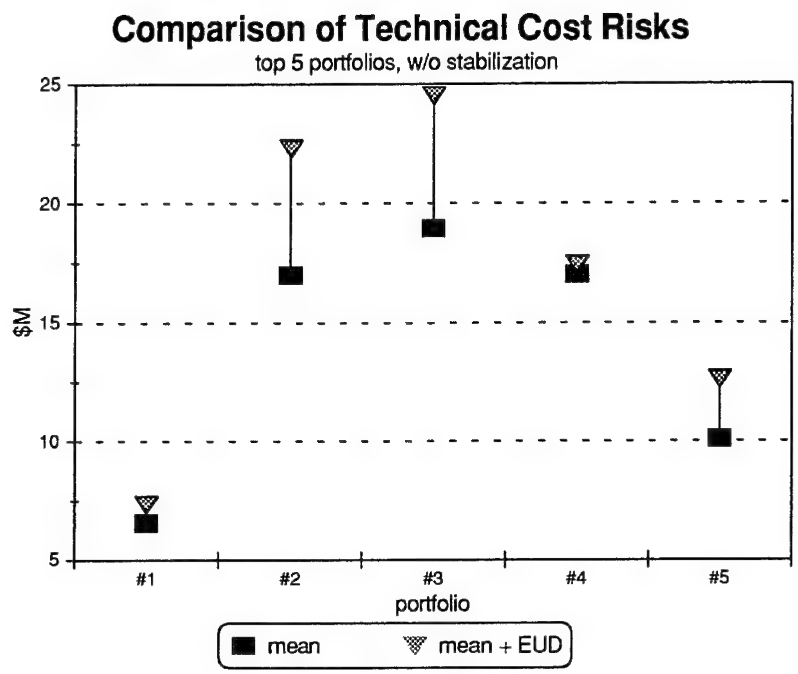


Figure 4.21

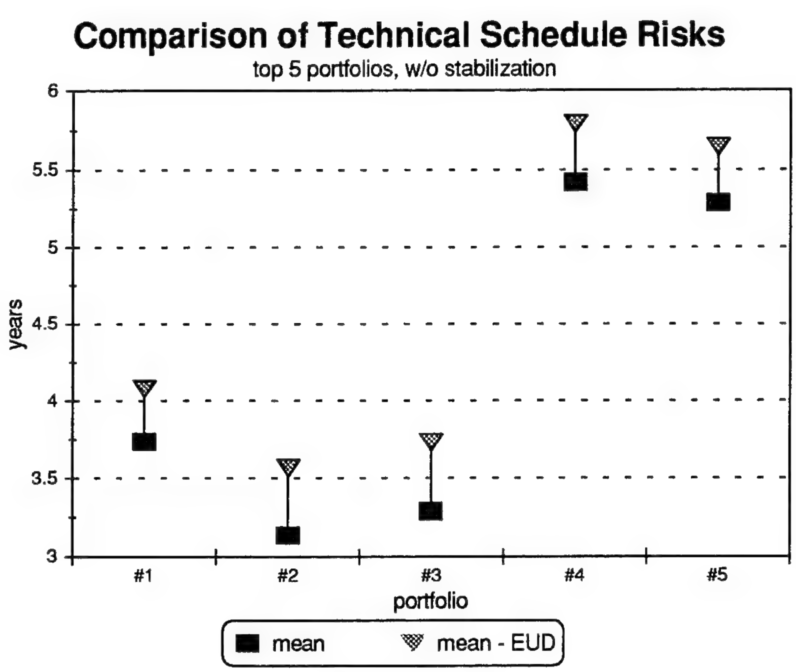


Figure 4.22

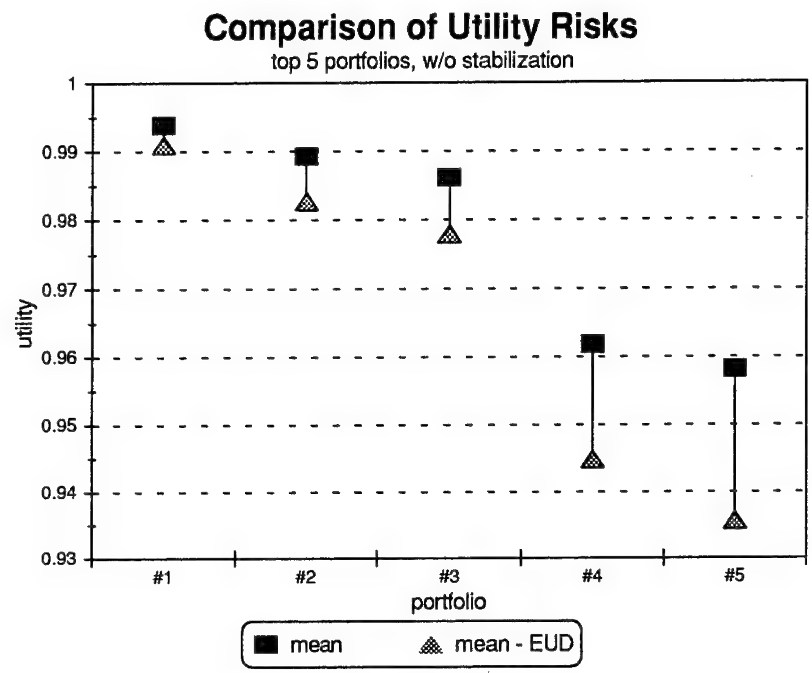


Figure 4.23

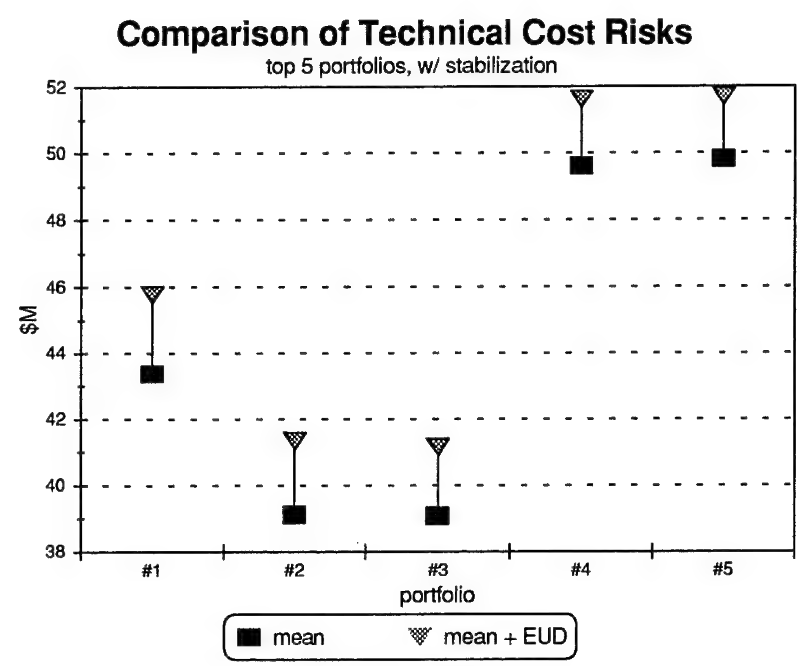


Figure 4.24

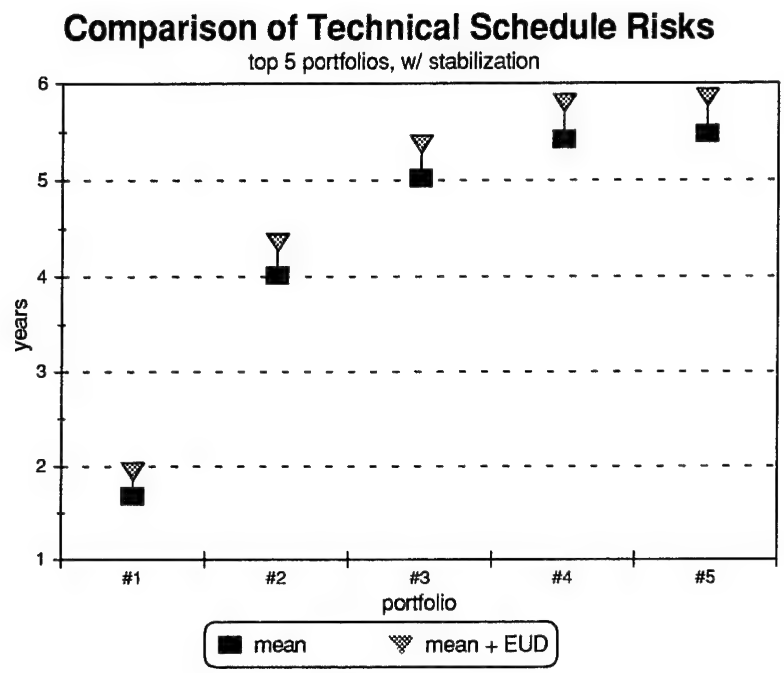


Figure 4.25

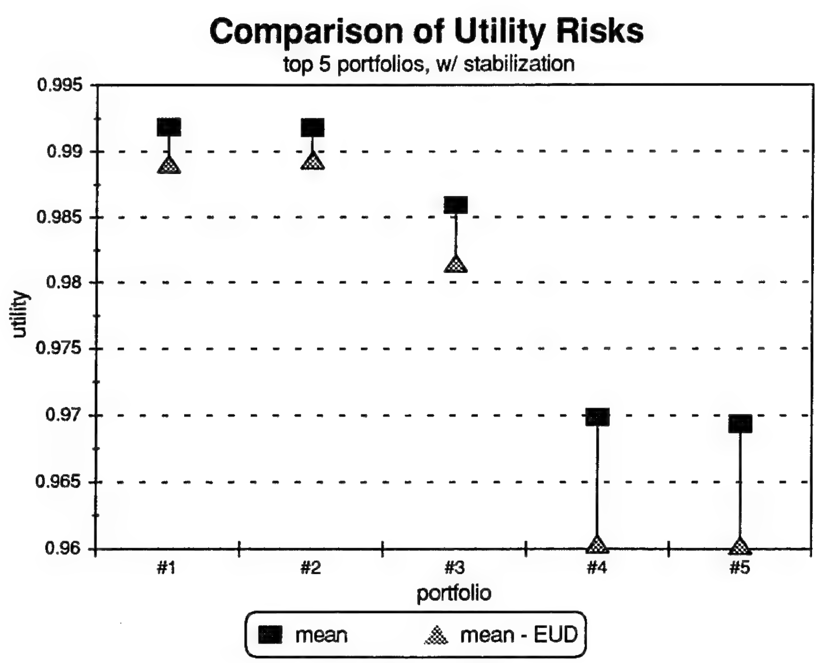


Figure 4.26

Turning our attention to the portfolios employing stabilization, the risks seem to be relatively constant for all five. Choosing between portfolio #1 (means of $43.37M and 1.68 years) and #2 ($39.11M and 4.01 years) hinges on the decision maker's trade-off between cost and completion time — if a lower cost is favored more than a shorter remediation schedule, #2 would be the best choice, while #1 is preferred if the counter is true. This is multi-attribute utility theory's greatest contribution. It quantifies the decision maker's preferences for trading off the important decision factors. Figure 4.26 shows how close the total utility scores (means) are with the current weights. Notice that #2 actually has an EUD slightly less than the #1, the only case of utility EUD being smaller for a lower ranked alternative in this example data set. This EUD is dependent on the relative weighting between cost and schedule, as well, making interpretation difficult. But with the current weights, this lower EUD may make #2 more attractive to a decision maker than the slightly higher utility score of #1.

These graphs (Figures 4.9-4.26) summarize the cost and schedule risks in a concise and clear fashion. Both parts of risk — unfavorable consequence and probability — are represented by the length of the expected deviation line extending above the mean value. These cost and time expected unfavorable deviations are independent from the value assessed by utility functions and so represent additional decision-making criteria that can be used as needed to distinguish between alternatives. The EUDs of the utilities provide a sense of the utility PDFs of these alternatives, providing more information than just the expected utilities alone.

Means and EUDs For the Top Ten Portfolios

	Cost		Time		Total Utility	
portfolio	mean ($M)	EUD ($M)	mean (years)	EUD (years)	mean (utility)	EUD (utility)
without stabilization						
#1	6.56	0.76	3.73	0.35	0.99379	0.00286
#2	16.98	5.33	3.14	0.43	0.98926	0.00657
#3	18.94	5.58	3.29	0.44	0.98615	0.00826
#4	17.01	0.4	5.42	0.37	0.96184	0.01705
#5	10.07	2.55	5.29	0.36	0.95822	0.02257
with stabilization						
#1	43.37	2.4	1.68	0.27	0.99184	0.00277
#2	39.11	2.23	4.01	0.35	0.9918	0.00243
#3	39.08	2.08	5.02	0.35	0.98589	0.00447
#4	49.6	2.03	5.43	0.37	0.96986	0.00951
#5	49.81	1.89	5.48	0.37	0.96935	0.00914

Table 4.2

4.2.4 *Semi-variances and Coefficients of Variation.* Table 4.3 shows the variances and semi-variances of the top ten alternatives.

Figures 4.27-4.32 show the variances and semi-variances compared against the EUDs as measures of risk. The heights of the bars reflect the magnitude of that risk measure for that alternative, and so the rankings of each alternative by risk measure can

Variances and Semi-variances For the Top Ten Portfolios

	Cost		Time		Total Utility	
portfolio	variance ($M^2)	semi-variance ($M)	variance (years^2)	semi-variance (years^2)	variance (utility^2)	semi-variance (utility^2)
without stabilization						
#1	3.9106	2.3646	0.8205	0.4835	0.00014	0.00013
#2	197.63	162.23	1.4171	1.0087	0.00055	0.0006
#3	205.63	164.95	1.4322	1.0103	0.00105	0.00096
#4	1.4657	0.7294	0.9139	0.5685	0.00353	0.00031
#5	82.688	75.311	1.1885	0.7838	0.00674	0.0061
with stabilization						
#1	47.873	31.599	0.47119	0.32545	0.00016	0.00014
#2	39.806	26.148	0.85586	0.50475	0.00007	0.00006
#3	35.0806	23.062	0.8802	0.51947	0.00021	0.00024
#4	37.215	24.835	0.91076	0.56713	0.00236	0.00221
#5	33.281	22.096	0.90917	0.56642	0.00217	0.00202

Table 4.3

be determined. Since EUD is in different units than the variance and semi-variance, it is plotted against the left axis instead of the right. Of particular interest are those cases where the rankings would be different based on variance and semi-variance, and EUD and semi-variance. Again, care should be taken when interpreting the risk measures of the utility scores.

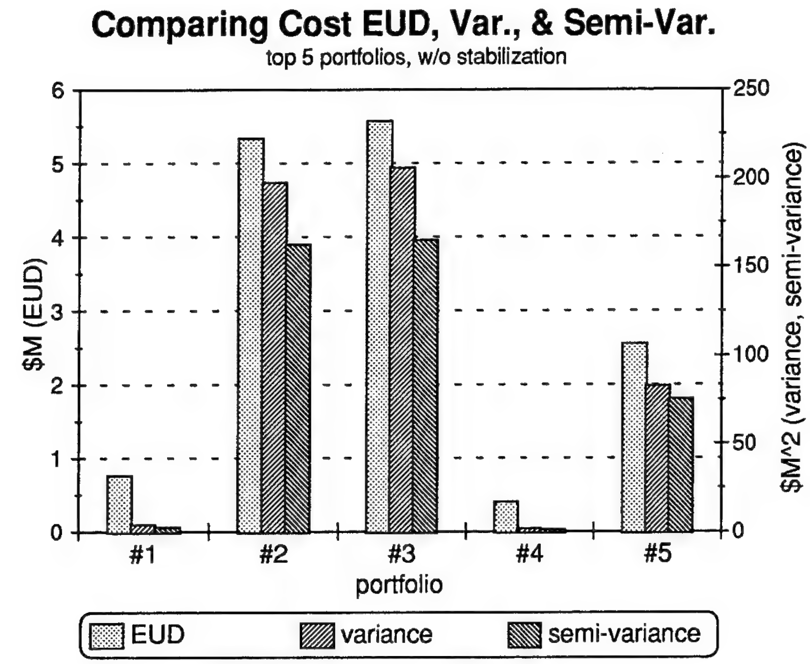


Figure 4.27

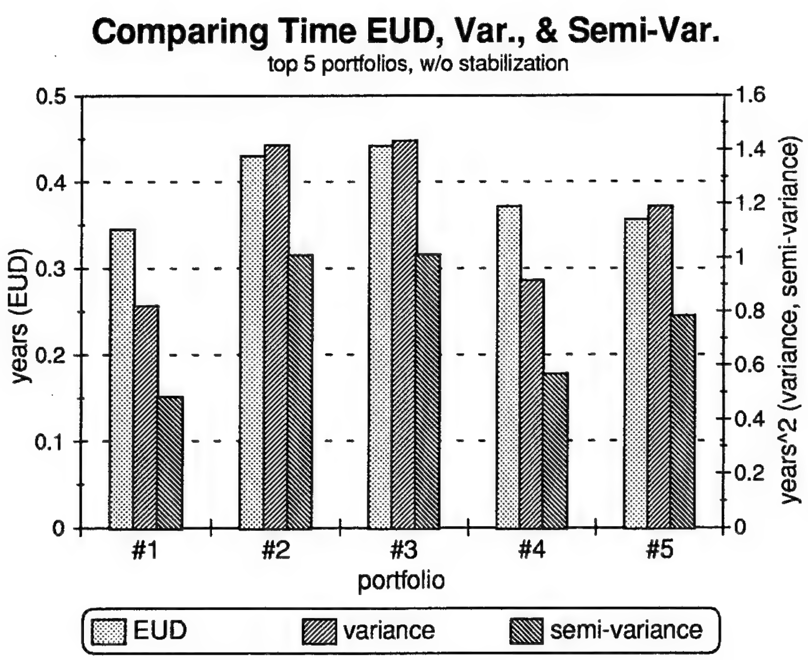


Figure 4.28

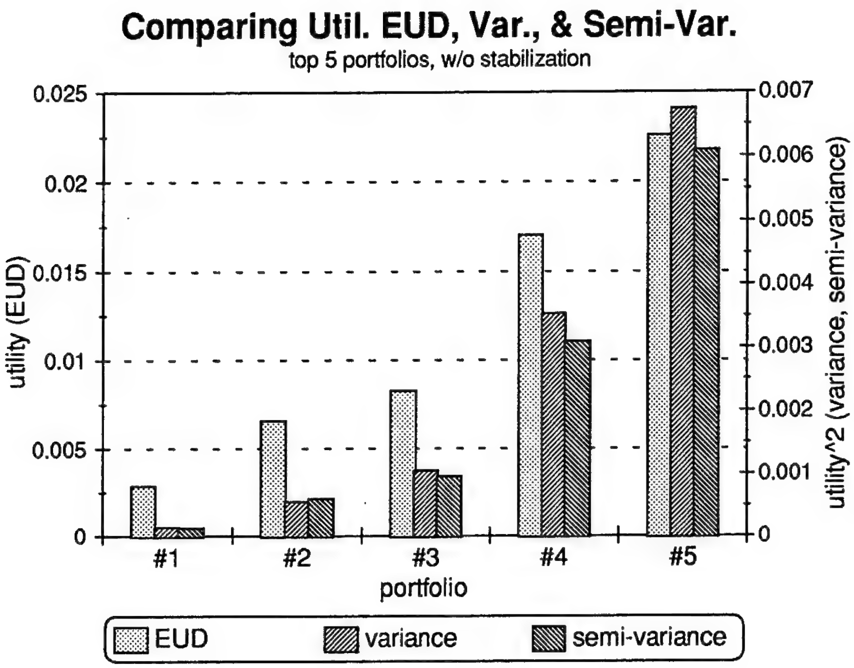


Figure 4.29

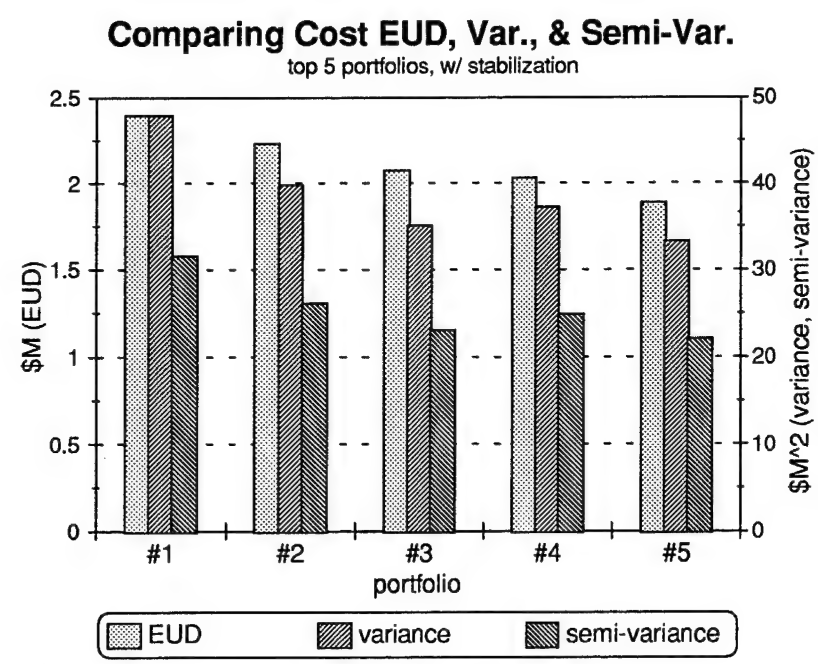


Figure 4.30

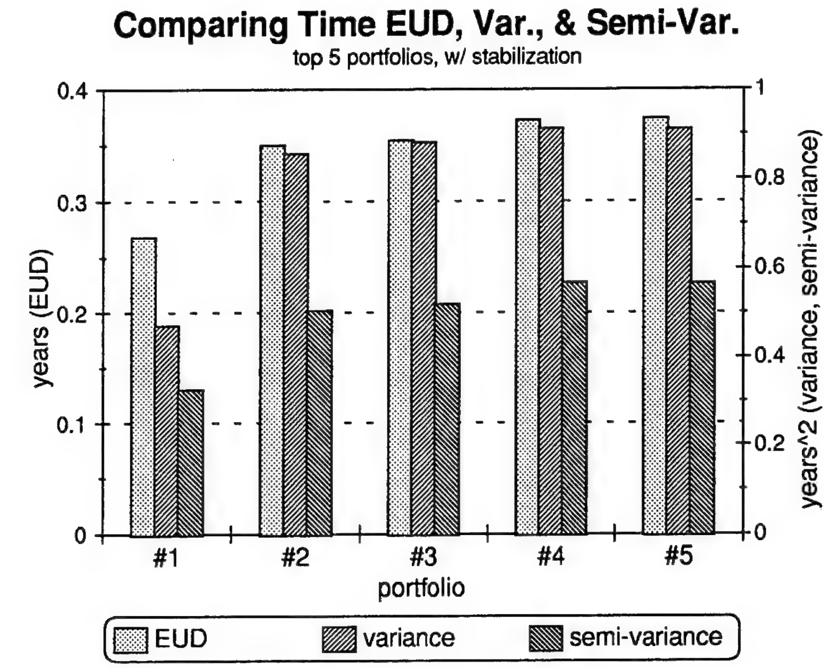


Figure 4.31

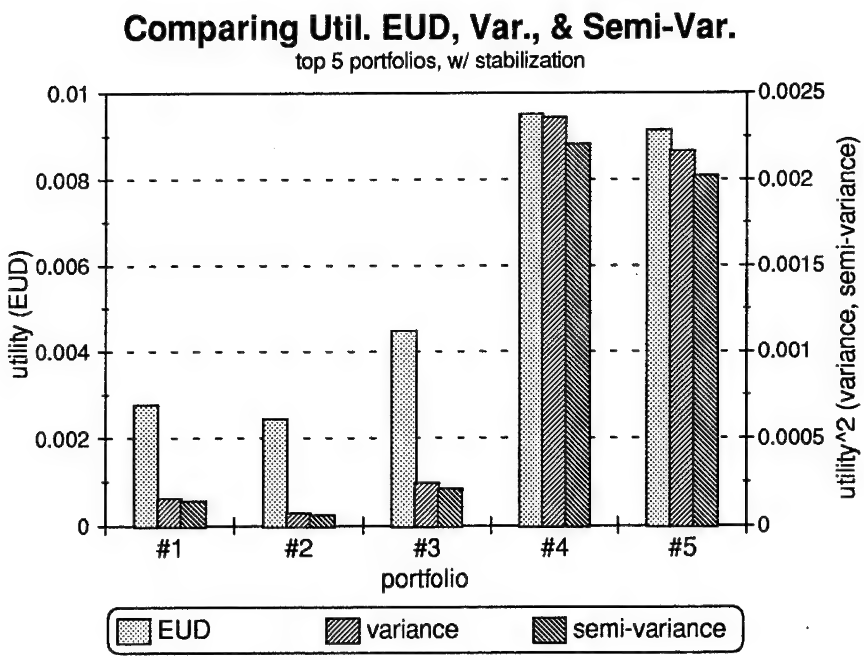


Figure 4.32

Figure 4.28 shows one situation where using semi-variance would result in a different ranking by risk than using EUD. Here, looking at the schedule risk measures for the non-stabilized portfolios, the three least risky portfolios are (in order of decreasing risk) #4-#5-#1 for EUD and #5-#4-#1 for semi-variance (and variance, as well). Another examples of different rank ordering can be seen on Figure 4.30, where the cost risk measures for the stabilized portfolios result in swapped third and fourth most risky positions: EUD results in #3-#4 while semi-variance and variance result in #4-#3. This confirms the discussion in section 3.4.4.2 in Chapter III.

The coefficient of variation, the standard deviation divided by the mean, is suggested by finance references as a measure of relative risk [VanHorne, 1971:46]. The coefficient of variations of the ten portfolios are shown in Table 4.4 and Figures 4.24-25.

Coefficients of Variation

portfolio	#1	#2	#3	#4	#5
non-stabilized					
cost	0.3013	0.8277	0.7577	0.0712	0.9027
time	0.2426	0.3797	0.364	0.1764	0.2062
utility	0.012	0.0236	0.0329	0.0618	0.0857
stabilized					
cost	0.1595	0.1613	0.1516	0.123	0.1158
time	0.4084	0.2305	0.1868	0.1759	0.1741
utility	0.0125	0.0086	0.0158	0.0501	0.048

Table 4.4

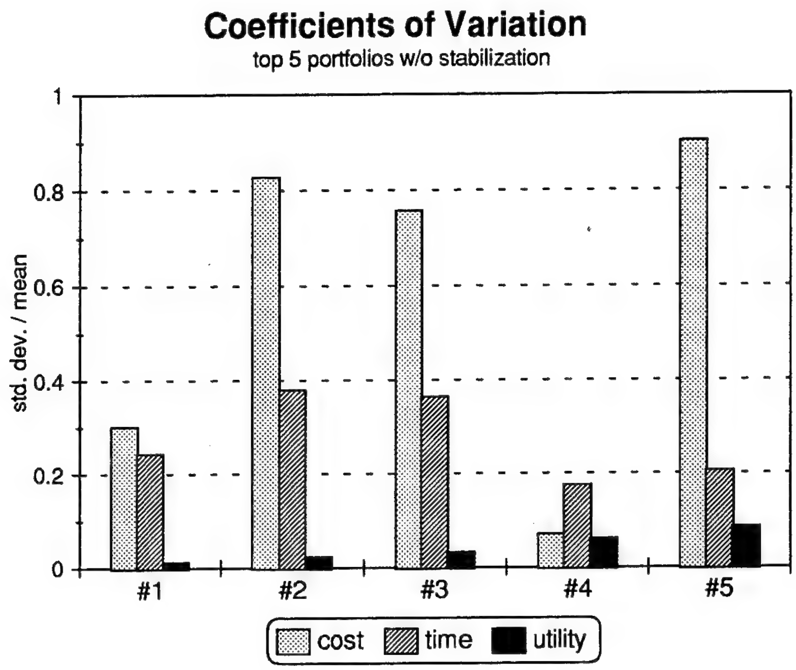


Figure 4.33

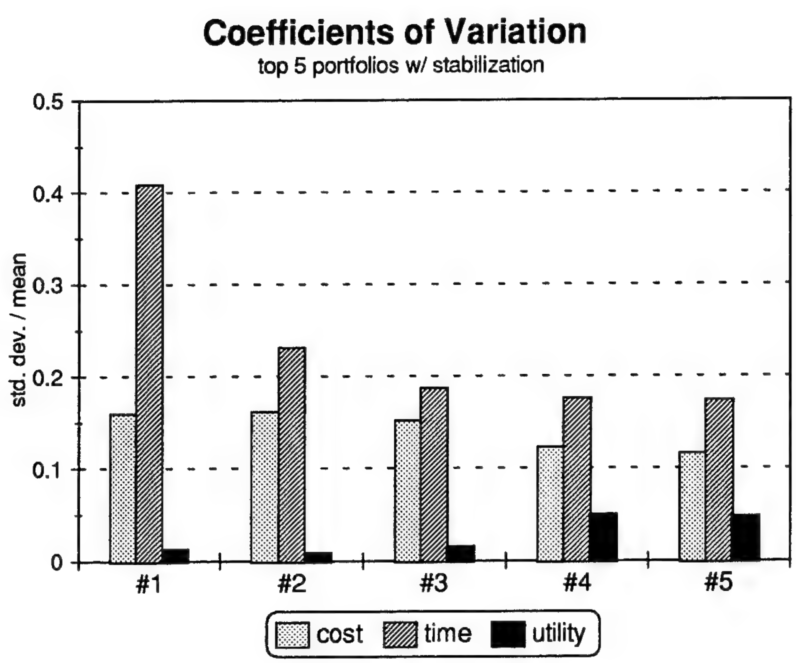


Figure 4.34

Normalized EUDs

portfolio	#1	#2	#3	#4	#5
non-stabilized					
cost	0.1161	0.3141	0.2945	0.0237	0.2535
time	0.0925	0.1373	0.1343	0.0686	0.0673
utility	0.002877	0.006645	0.008373	0.017726	0.023559
stabilized					
cost	0.0552	0.0571	0.0531	0.0409	0.0379
time	0.1593	0.0872	0.0705	0.0686	0.0682
utility	0.002793	0.002453	0.004537	0.009801	0.009429

Table 4.5

Since the coefficient of variation is based on the variance, which is not an accurate measure of the unfavorable variation alone, they are not good measures of risk according to our definition. However, the EUDs can be normalized by the means as well to form a relative measure of risk as well. These EUDs divided by the means are shown in Table 4.5. Figures 4.35-4.40 display these "normalized" EUDs compared with the coefficient of variations in order to contrast risk rankings resulting from the relative heights of the bars.

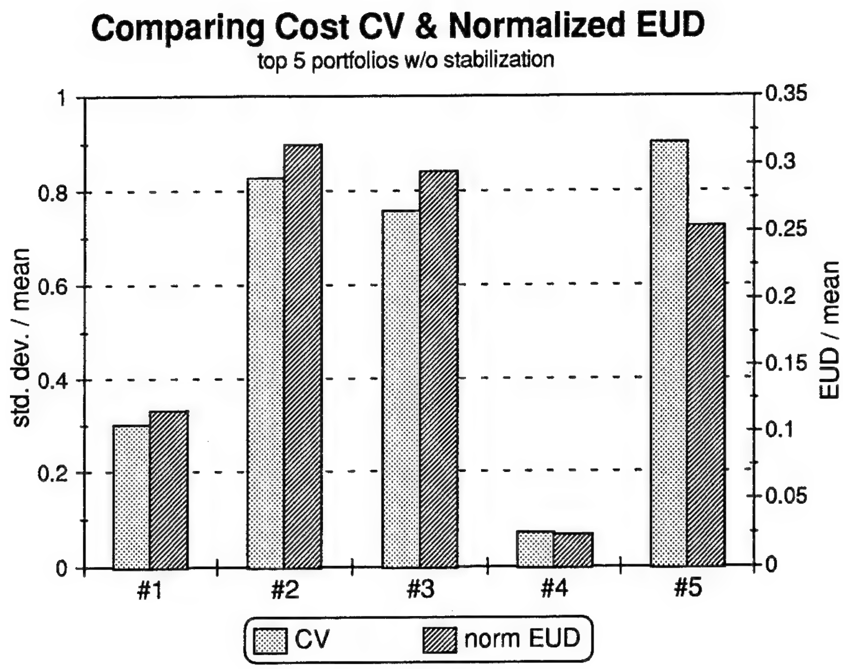


Figure 4.35

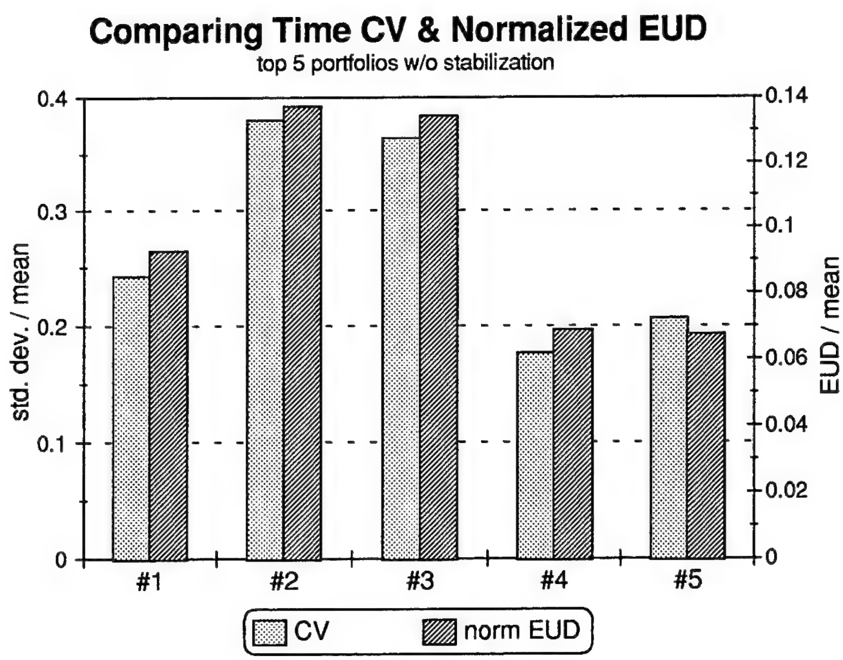


Figure 4.36

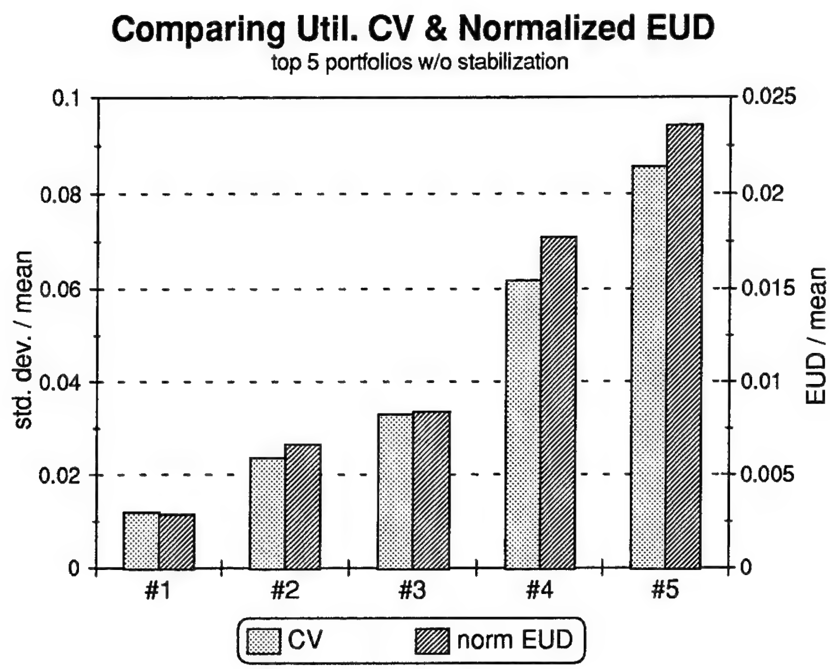


Figure 4.37

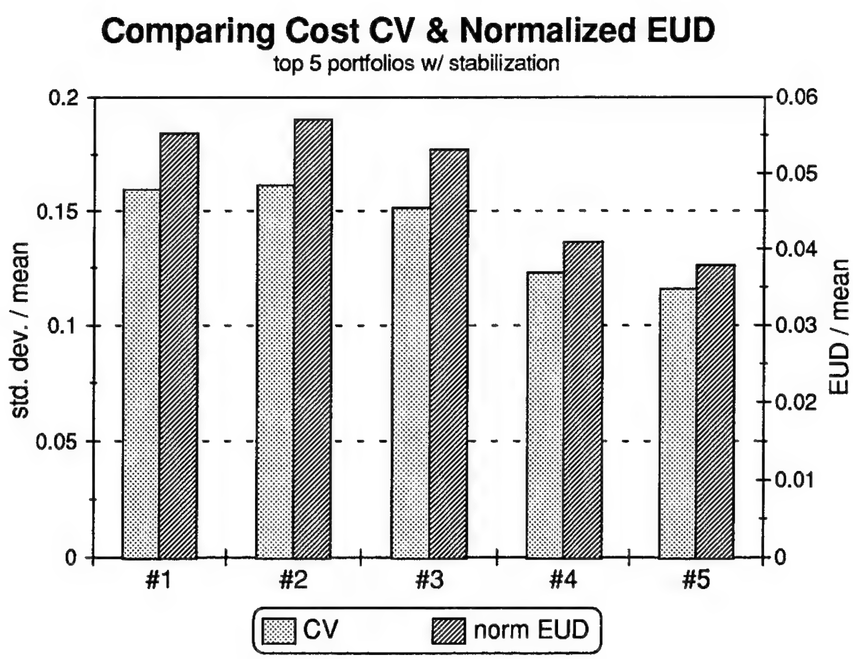


Figure 4.38

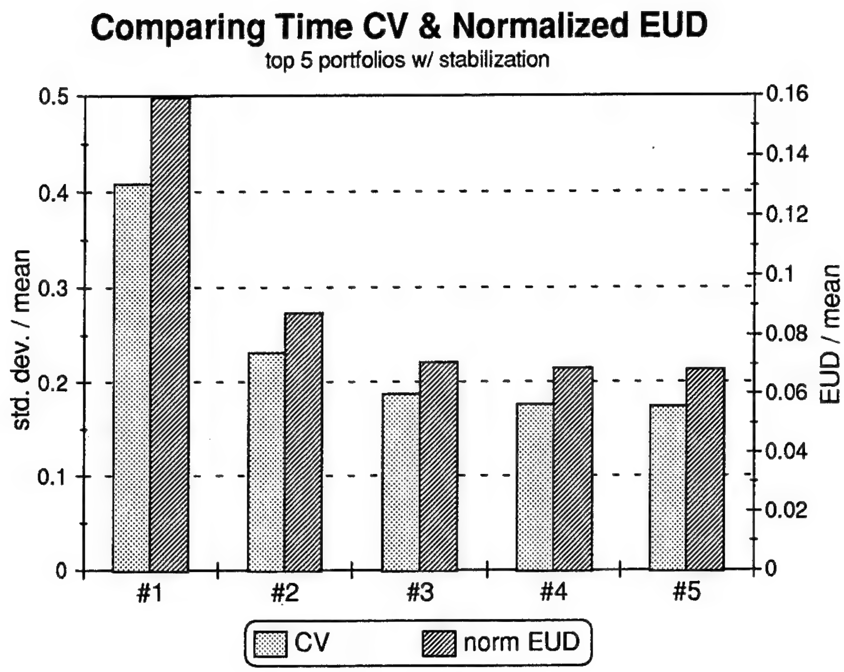


Figure 4.39

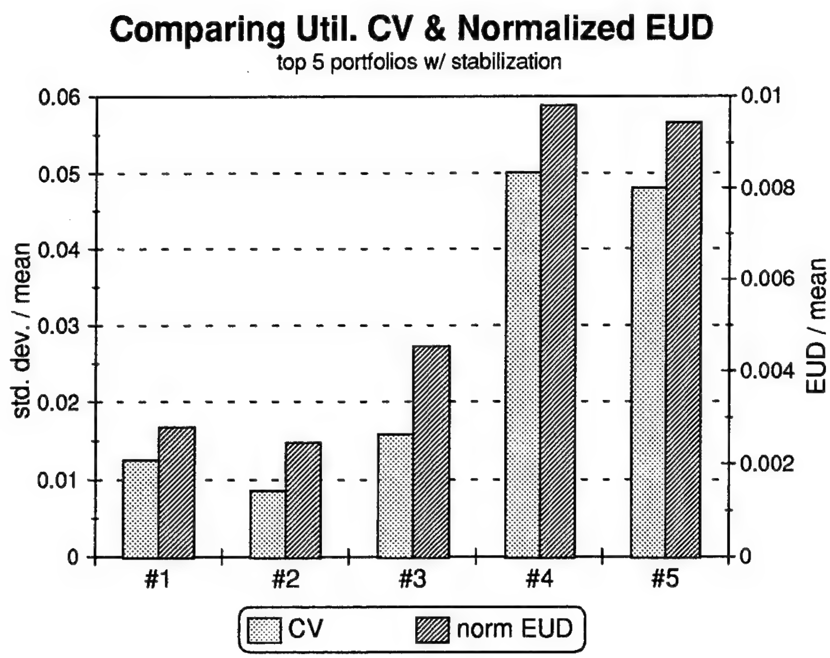


Figure 4.40

The two relative risk measure produce different rankings for the unstabilized portfolios, where the coefficient of variation yields a #5-#2-#3-#1-#4 order for cost and a #2-#3-#1-#5-#4 for time but the normalized EUD yields #2-#3-#5-#1-#4 and #2-#3-#1-#4-#5. The most interesting thing is the difference in risk ranking between the standard EUD and the normalized EUD, as summarized in Table 4.6.

The cost rankings are little different from the standard and the normalized EUDs. Only the stabilized #3 and #2 swapped places, and they have scores that are close together in both measures. The time rankings, however, show surprising changes for all portfolios. The complete reversal in rankings for the stabilized portfolios makes more

Risk Rankings for EUD and Normalized EUD

from most to least risky	for cost	for time
non-stabilized portfolios		
ranked by EUD	#3-#2-#5-#1-#4	#3-#2-#4-#5-#1
ranked by norm. EUD	#2-#3-#5-#1-#4	#2-#3-#1-#4-#5
stabilized portfolios		
ranked by EUD	#1-#2-#3-#4-#5	#5-#4-#3-#2-#1
ranked by norm. EUD	#1-#2-#3-#4-#5	#1-#2-#3-#4-#5

Table 4.6

sense when the magnitude of the EUDs are examined in Figure 4.25, as they are all relatively the same. The difference in means (see Figure 4.24) then dominates. Similar effects are causing the swapping of position in the non-stabilized time rankings.

The semi-variance could be used in place of the variance, to form a "coefficient of semi-variance." This would measure the relative downside risk in a similar fashion as the

normalized EUD, with the same difficulties when the deviation from the mean is less than 1.

The coefficient of variation and normalized EUD are relative risk measures, but by dividing by the mean, the risk expressed solely by the shape of the variables' distributions is confounded with a measure of value. They are unitless quantities, and therefore may not have much meaning to a program manager who wants to know the actual dollar or year risk.

4.2.5 *Summary of Risk Measures*. We have examined many ways of quantifying risk. By breaking the objective cost and schedule distributions out from the subjective utility scores, we can give the decision maker much more information that will impact his or her decisions. The range graphs, showing the bounds and expected value of our output PDFs, show the potential best, worst, and most likely cases for each portfolio. When combined with the mean + EUD charts, these graphs convey the cost and schedule risks of each portfolio in a concise and easy-to-understand manner. We compared the EUD measure of risk to variance and semi-variance, and found that with our notional data they would generate different risk rankings. This makes EUD more attractive than semi-variance, because of the problems with squaring deviations that are less than one. Relative risk measures such as the coefficient of variation and the similar normalized EUD resulted in different rankings in some portfolios as well, but their usefulness as unitless quantities to a practical decision maker concerned about dollars and schedule months is debatable.

4.2.6 *Sensitivity to Estimates of the Probability of Successful Implementation.* The recommendations of the Decision Analysis Module (i.e. technology portfolio selection) may be sensitive to changes in the estimates of P(use). If errors in P(use) have a large effect on the results, the recommendations of the decision support system could be subject to dispute. It would be necessary then to more accurately determine the P(use) parameter. However, it may be difficult to increase the accuracy of the P(use) estimates, as discussed in Chapter III, section 3.3.4.

To examine the sensitivity of the preliminary results to changes in P(use), two additional cases were examined in detail for four technology portfolios. The levels of P(use) were raised by 10% (to a maximum of 100%) for all of the portfolio's technologies and the effects quantified. The same portfolios then had their P(use) lowered by 10% (to a minimum of 0). This way potential systematic over- and underestimations could be examined. While these are not the most stressing cases of potential mis-assessment, some idea of the potential effects can be gained. The #1 and #3 portfolios for both the non-stabilized and stabilized strategies were examined to illustrate this concept. These four were chosen to cover both retrieval-treatment-disposal and containment strategies for the non-stabilized case, and to check more than one stabilized portfolio. A more complete examination of the sensitivity to P(use) should be accomplished when analyzing actual sponsor-donated data with a fully running LCC model.

4.2.6.1 *Graphical Comparisons.* Figures 4.41-4.60 show the different range and EUD graphs for these four portfolios.

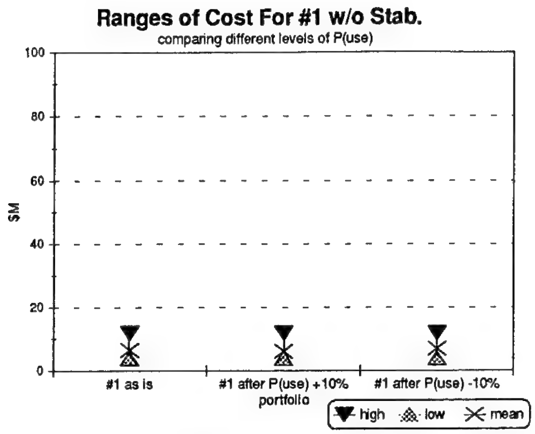


Figure 4.41

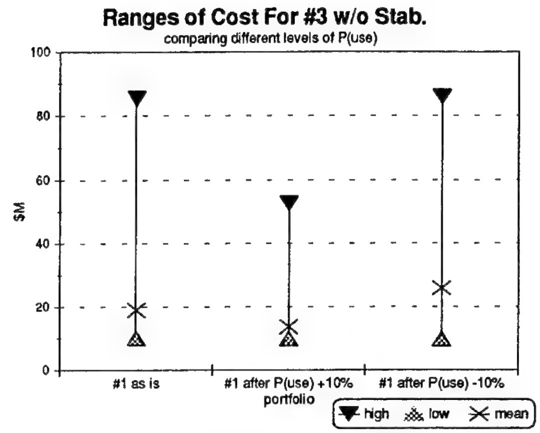


Figure 4.42

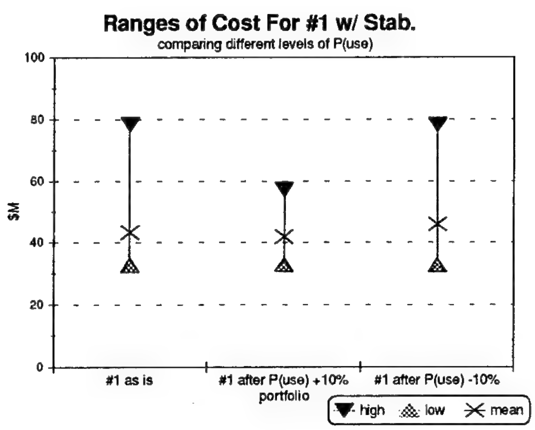


Figure 4.43

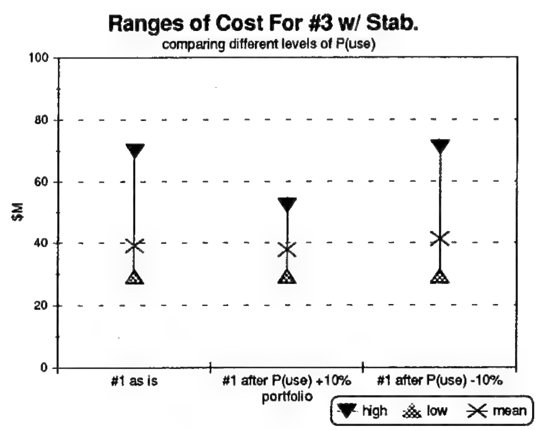


Figure 4.44

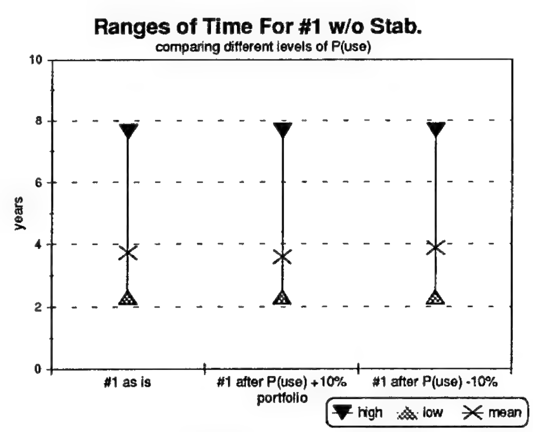


Figure 4.45

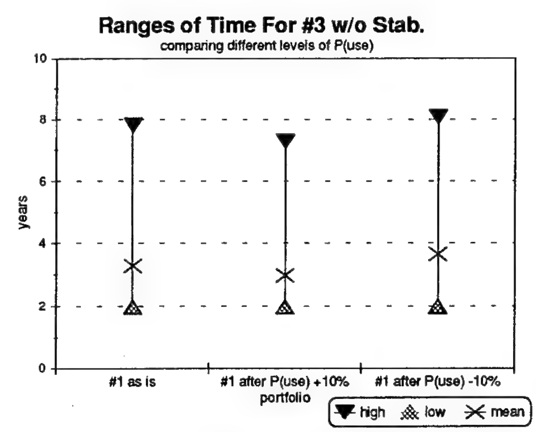


Figure 4.46

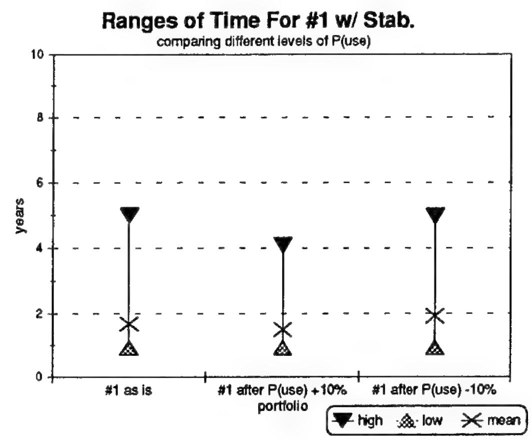


Figure 4.47

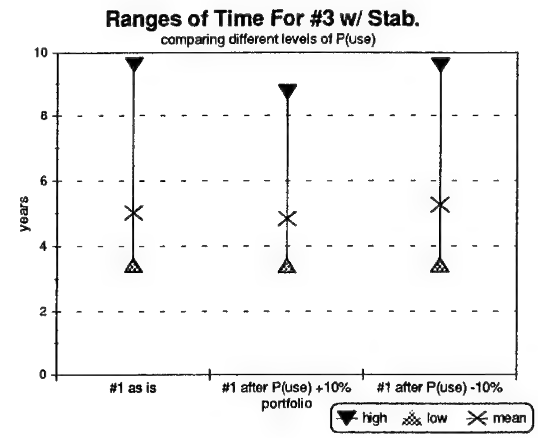


Figure 4.48

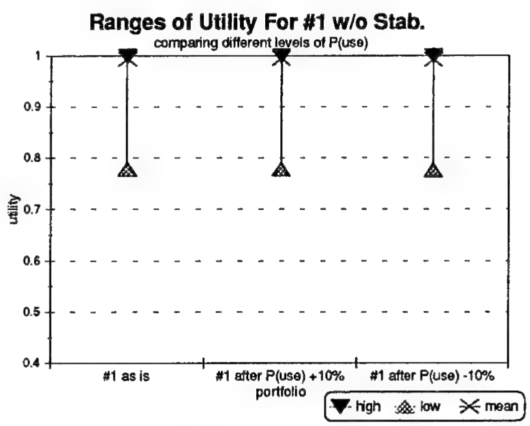


Figure 4.49

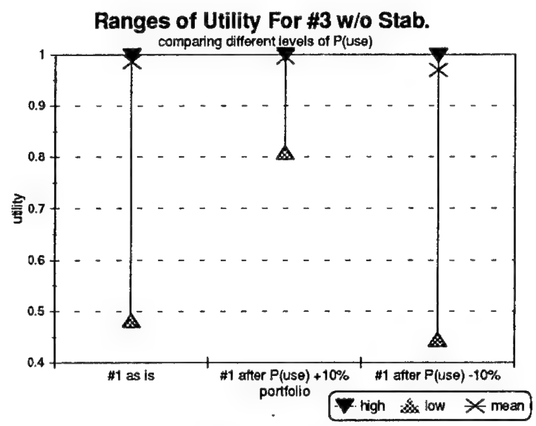


Figure 4.50

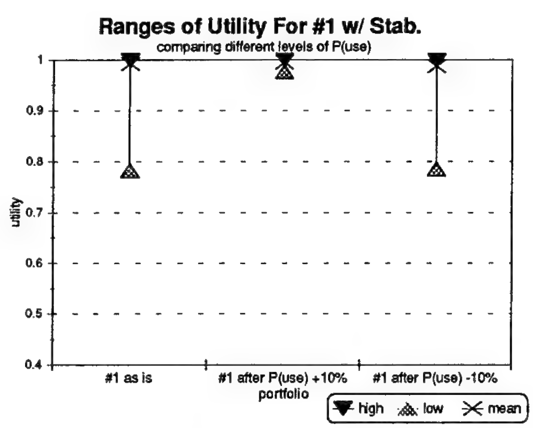


Figure 4.51

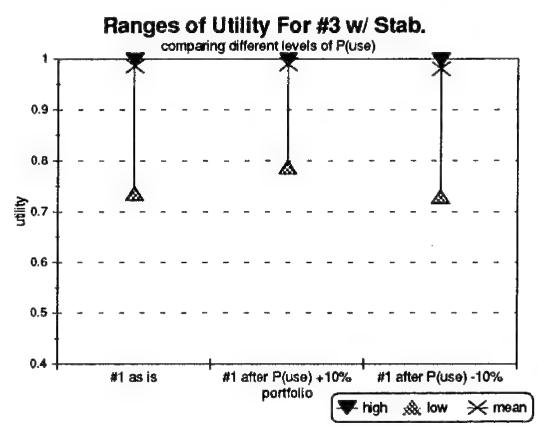


Figure 4.52

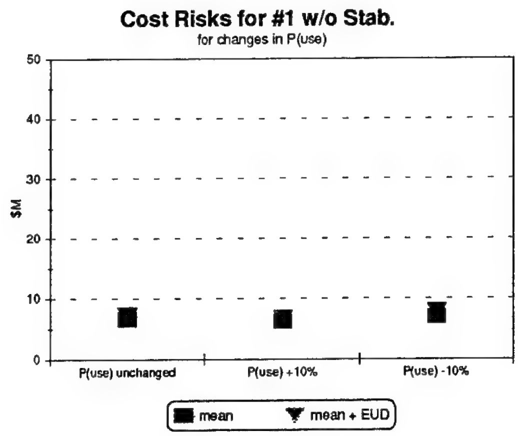


Figure 4.53

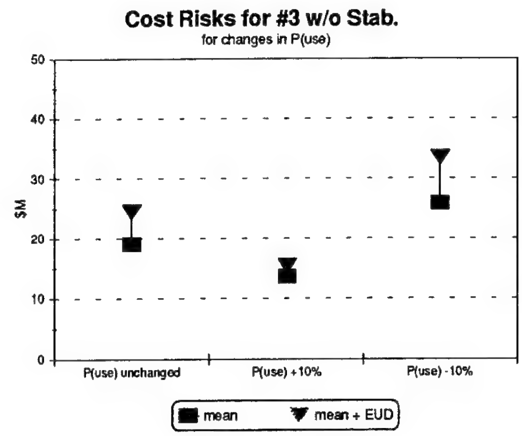


Figure 4.54

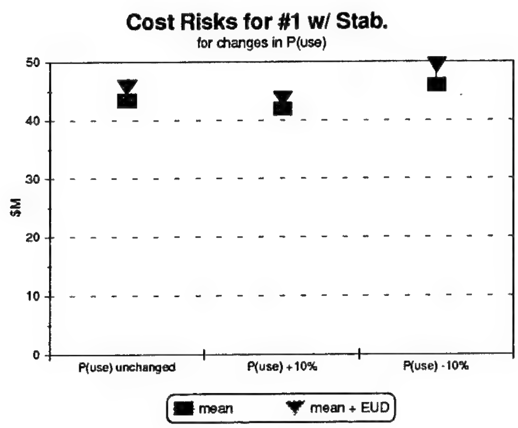


Figure 4.55

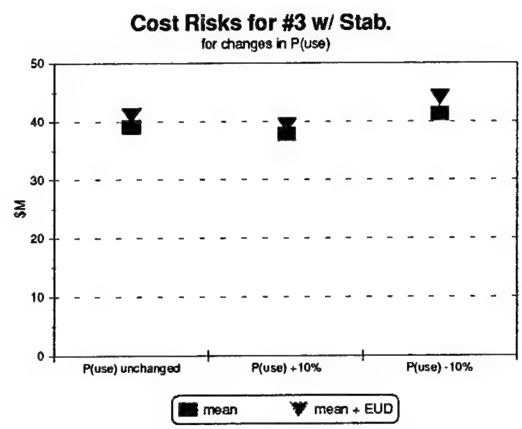


Figure 4.56

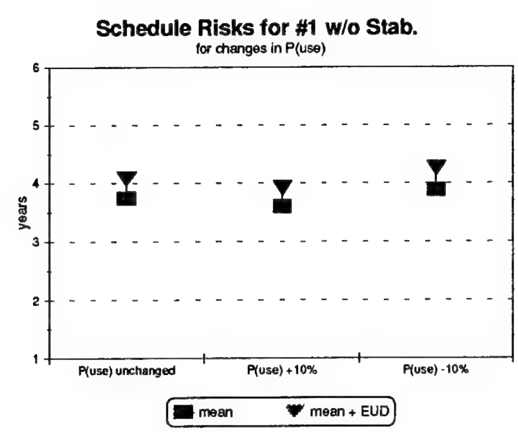


Figure 4.57

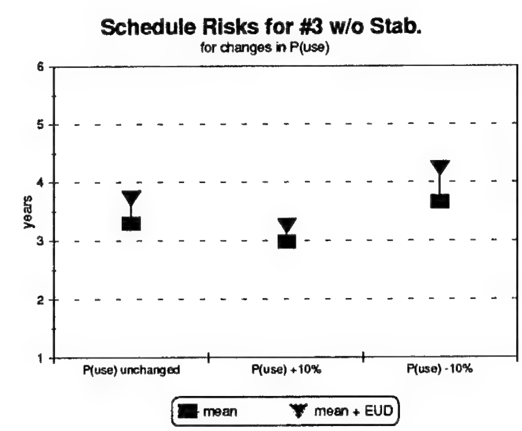


Figure 4.58

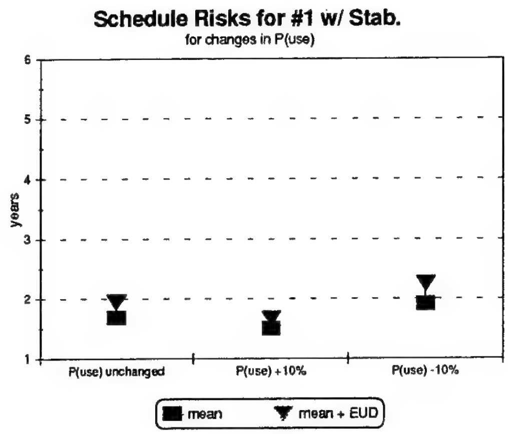


Figure 4.59

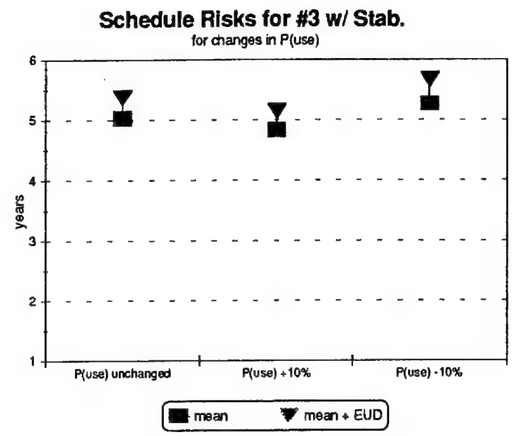


Figure 4.60

Examination of the graphs of ranges in Figures 4.41 to 4.48 shows no great effect of lowering P(use) by 10% for all technologies. The mean costs and times rise slightly, but the high costs and times remain mostly the same. Raising P(use) lowers the probabilities of the highest costs and times, as one would expect from lower chances of incurring the penalty times and costs. Consequently, the probabilities of the lowest utilities change as well. The graphs of utility ranges, Figures 4.49-4.52, show large changes in the lowest utilities for the unstabilized #3 and stabilized #1 portfolios, a small change in the low point for stabilized #3, and little or no change for unstabilized #1. In general, the ranges of time remained fairly constant while increasing P(use) dramatically lowered the highest costs for all but unstabilized #1.

More effects can be seen on the graphs of cost and schedule means and EUDs, Figures 4.53-60. The unstabilized #3 portfolio in particular has a shift in mean cost as P(use) is raised (mean drops from \$18.94M to \$13.54M) and lowered (mean rises to \$25.77M; see Figure 4.39). There was little change in schedule risk as P(use) changed in

Figures 4.48-4.51. In general, risk increases when P(use) is lowered and decreases when P(use) is increased.

4.2.6.2 *Statistical Testing.* To confirm the conclusions drawn from the graphs, statistical tests of hypotheses were used to examine the impact of the systematic changes in P(use). The simulation results were treated as samples drawn from the population that would have resulted from the use of full enumeration in the DA model. First, the variances of the basecase were compared to the raised P(use) results and the lowered P(use) results to see how different they were. This procedure is summarized in Table 4.7 below. Then, the means of the results were compared to see if they were statistically different, using the procedure in Table 4.9.

Test of Equal Variances

H_o: $\sigma_1^2 = \sigma_2^2$

H_a: $\sigma_1^2 \neq \sigma_2^2$

Test Statistic: $\hat{F} = \dfrac{\max\ (S_1^2,\ S_2^2)}{\min\ (S_1^2,\ S_2^2)}$

RR: $\hat{F} > F_{\frac{\alpha}{2},\ n_L - 1,\ n_H - 1}$

where n_H corresponds to the largest S^2 and n_L to the smallest

Assumptions: Two samples are independent and normally distributed.

Table 4.7 [Mendenhall, et. al., 1990:468-9]

The normality assumptions provide some difficulty, but with 10,000 samples and some caution this test can still be applied. There was some difficulty in finding the

rejection region, since most tables or software for the F distribution do not reach degrees of freedom as high as 10,000/10,000 before going to the limit at infinity. However, we can bound the appropriate F statistic since we know

$$F_{\frac{\alpha}{2},\ 1000,\ 1000} > F_{\frac{\alpha}{2},\ 9999,\ 9999} > F_{\frac{\alpha}{2},\ \infty,\ \infty}. \qquad (4.2)$$

and $F_{\frac{\alpha}{2},\ \infty,\ \infty} = 1$ for all α. Therefore, if the test statistic $\hat{F} > F_{\frac{\alpha}{2},\ 1000,\ 1000}$, we know for certain that we can reject the null hypothesis for that significance level α. If $\hat{F} < F_{\frac{\alpha}{2},\ 1000,\ 1000}$, on the other hand, we cannot say for certain that we fail to reject H_o since the true rejection region threshold is less than $F_{\frac{\alpha}{2},\ 1000,\ 1000}$. With this in mind, Table 4.8 shows the necessary significance level α required for the test statistic $\frac{\max\ (S_1^2,\ S_2^2)}{\min\ (S_1^2,\ S_2^2)} >$ $F_{\frac{\alpha}{2},\ 1000,\ 1000}$.

As these significance levels show, at an α of 0.01 we can reject the null hypothesis in all but one case, that of the completion time of the #1 non-stabilized portfolio when lowering P(use). Since the actual rejection region threshold is lower than that used for the above table, that case may still reflect different population variances. In general, we can say with high confidence (1 - α) that changing P(use) had a statistically significant effect on the variance of the output cost, time, and utility distributions, if the normality assumption was justified. Although we cannot accept this normality assumption, we can cautiously say that the systematic changes in P(use) had a demonstrable effect on the variance of the results.

Results of Testing Equal Variances

		Cost		Time		Total Utility	
		$\hat{F}$	α	$\hat{F}$	α	$\hat{F}$	α
unstab. #1	P(use)+	1.302	0.00001	1.208	0.00071	1.68	$< 5 \times 10^{-6}$
	P(use)-	1.157	0.005305	1.138	0.01027	1.453	$< 5 \times 10^{-6}$
unstab. #3	P(use)+	3.074	$< 5 \times 10^{-6}$	1.622	$< 5 \times 10^{-6}$	3.638	$< 5 \times 10^{-6}$
	P(use)-	1.577	$< 5 \times 10^{-6}$	1.388	$< 5 \times 10^{-6}$	3.55	$< 5 \times 10^{-6}$
stab. #1	P(use)+	1.898	$< 5 \times 10^{-6}$	1.629	$< 5 \times 10^{-6}$	14.131	$< 5 \times 10^{-6}$
	P(use)-	1.787	$< 5 \times 10^{-6}$	1.382	$< 5 \times 10^{-6}$	2.878	$< 5 \times 10^{-6}$
stab. #3	P(use)+	1.872	$< 5 \times 10^{-6}$	1.291	0.000015	2.21	$< 5 \times 10^{-6}$
	P(use)-	1.792	$< 5 \times 10^{-6}$	1.179	0.002345	2.15	$< 5 \times 10^{-6}$

Table 4.8

Since we know that $\sigma_1^2 \neq \sigma_2^2$, testing to see if the difference between the means of the basecase and the changed cases becomes difficult. Classical hypothesis tests do not cover this situation. However, Law and Kelton do describe an approximation that allows one to make confidence intervals on the difference of two means from approximately normal distributions with unequal variances [1991:589]. Using this Welch approximation in a hypothesis test gives us the procedure in Table 4.9.

Test of Equal Means

H_o: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$

Test Statistic: $\hat{t} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

RR: $|\hat{t}| > t_{\frac{\alpha}{2},\hat{f}}$

$$\text{where } \hat{f} = \frac{(\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2})^2}{(\frac{s_1^2}{n_1})^2(n_1 - 1) + (\frac{s_2^2}{n_2})^2(n_2 - 1)}$$

Assumptions: Two samples are independent and normally distributed.

[Mendenhall, et. al., 1990:457; Law and Kelton, 1990:589]

Table 4.9

In our case of $n_1 = n_2 = 10{,}000$, the approximate degrees of freedom for the t statistic, $\hat{f}$, is approximately ∞, resulting in $t = 2.576$ for $\alpha = 0.01$. Table 4.10 below gives the results of this testing.

Again at the 99% significance level, we can say that changing P(use) had a statistically significant effect on the means of the output cost, time, and utility distributions, if the normality assumption was justified. Again although we cannot accept this assumption, we can cautiously say that the systematic changes in P(use) had a demonstrable effect.

Results of Testing Equal Means

		Cost		Time		Total Utility	
		$\hat{t}$	Result	$\hat{t}$	Result	$\hat{t}$	Result
unstab. #1	P(use)+	14.58	reject	11.06	reject	9.26	reject
	P(use)-	12.31	reject	10.75	reject	7.69	reject
unstab. #3	P(use)+	32.67	reject	20.63	reject	24.24	reject
	P(use)-	29.63	reject	19.42	reject	23.69	reject
stab. #1	P(use)+	17.15	reject	20.97	reject	20.39	reject
	P(use)-	21.81	reject	21.87	reject	20.9	reject
stab. #3	P(use)+	17.25	reject	15.26	reject	18.89	reject
	P(use)-	22.03	reject	16.68	reject	21.35	reject

Table 4.10

The statistical tests show that the changes in P(use) do have a statistically significant ($\alpha = .01$) effect on the resulting distributions — if these distributions are normally distributed. However, we know from the histograms that they are often highly skewed. The hypothesis test for the means being equal uses an approximation from Law and Kelton for use in generating confidence intervals, which they say are good approximations even if the actual distributions are not normal [1991:588]. This gives some justification for cautiously using the results of the statistical tests.

4.2.6.3 *Additional Portfolios.* While only these four portfolios were examined in detail, the other portfolios were also checked for the effects of systematic

changes in every P(use) estimate. Figures 4.61 and 4.62 show the different total utilities for each portfolio under all three P(use) conditions. As one can see from Figure 4.61, there is a case of rank order changing when P(use) is raised. The #5 portfolio is ranked

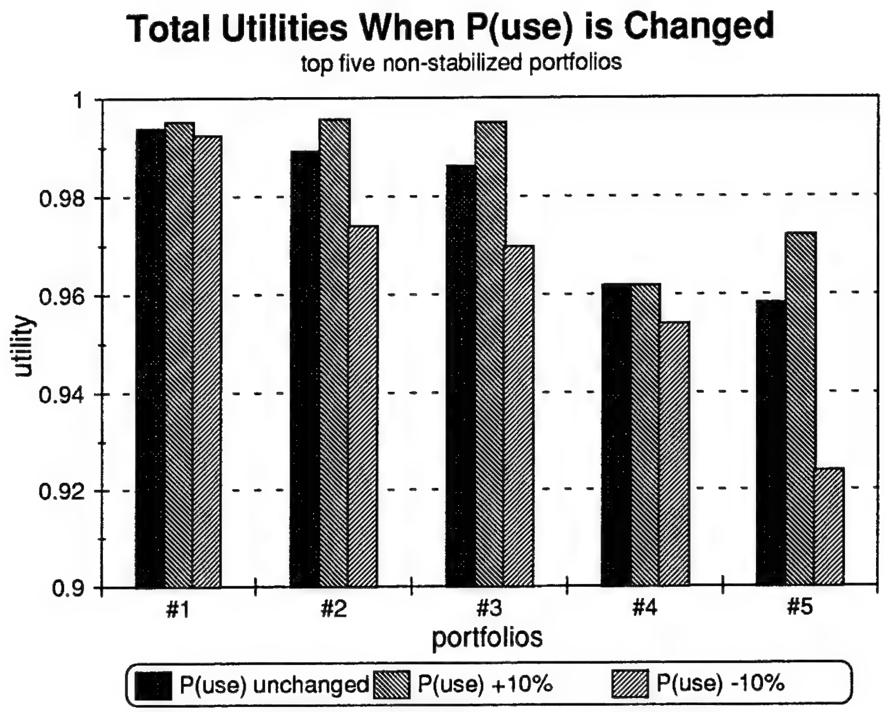


Figure 4.61

higher than the #4 one. In all other cases the relative rankings of these portfolios by total utility are the same.

Detailed sensitivity analysis can and should be done using the analysis tools that are part of the *DPL*© software to investigate the sensitivity of a recommendation to single values of P(use). In that way the criticality of individual assessments can be examined and

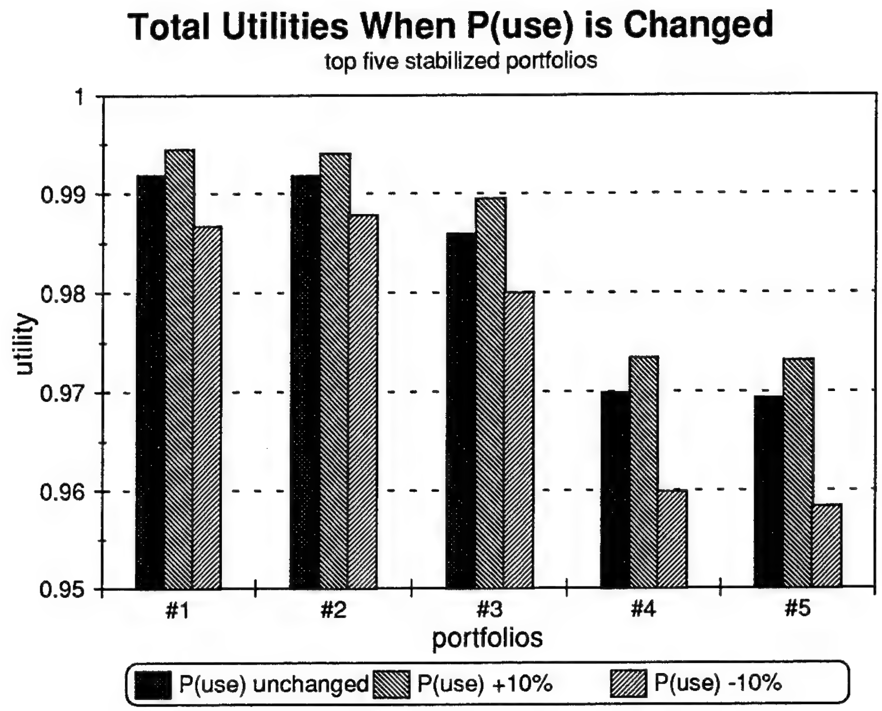


Figure 4.62

further investigated as needed. This is of great importance since we can see how changes in P(use) estimates may change the Decision Analysis Module's recommended technology portfolios.

V. *Conclusions and Recommendations*

5.1 *Conclusions From the Preliminary Study*

Working from the best engineering data available, we see trends developing from the results of the preliminary analysis described in Chapter IV. Containment strategies have much lower cost risks than retrieval-treatment-disposal strategies. Schedule risks are approximately the same for the top portfolios, leaving the mean required remediation time as the dominant discriminator between portfolios with this notional data. Including stabilization processes within a containment portfolio adds substantial cost and time. Some strategies (#4 and #5 with stabilization, #4 and #5 without stabilization) have the potential for unacceptable schedule overruns, with some costs near the $80M range despite mean costs of about $10-20M without stabilization and $30-50M with it. The model does not include the potential benefits of stabilizing the landfill, however, and safety and legal requirements may dictate the use of a stabilization strategy for specific sites.

These results may change when the Life-Cycle Cost Module is operational, since they are based on overall cost estimates for remediating 500,000 cubic feet of mixed waste instead of detailed models of the associated process. Still, containment strategies are likely to remain dramatically less cost risky than ex situ treatment strategies because the strategies are less complicated.

5.2 *Conclusions About the Methodology*

While subjective probability estimates have been used for technology selection [DOE, 1995e] and qualitative assessment of technical risk has played a role in evaluating different treatment technologies [Feizollahi and Quipp, 1995], quantifying the cost and schedule risks of candidate technology alternatives has not been done before for EM-50. The basic idea of Jia and Dyer, Weber, et. al., and others of quantifying risk using the variation about the expected value was applied through the simple expected unfavorable deviation (EUD) developed in Chapter III.

This independent measure of risk can be used as another decision criterion for each attribute, for risk averse decision makers. Mean cost and schedule results together with their EUDs can be used in a variety of ways to find the best technology strategy for a given application.

Subjective probability estimates for the duration of R&D, the likelihood of successful implementation, and the cost elements and capabilities of the LCC simulation model offer the best way to incorporate risk factors into the inputs of the decision support system. Risks of performance variability are then expressed through the measurable outputs of cost and time. These two attributes, total cost and overall schedule, are the two aspects of a remediation effort, apart from environmental and health risks, that are most important to our senior level decision makers. The final probability distributions that result from the Decision Analysis Module can then be condensed down to means, ranges, and EUDs with which we can compare alternatives.

While the preliminary study described here used R&D release date distribution that were originally estimated by experts using the earliest, most likely, and latest possible dates, several references have advocated soliciting opinions from the experts using the 10% and 90% fractiles instead of the absolute limits of the subjective distribution [Keefer and Bodily, 1983; Williams, 1994; Hudak, 1994]. This approach may limit the under-representation of the tails that motivated adjusting the distributions in Chapter III. While this may take additional explanation to solicit from experts, the results are worthwhile if the experts understand what is meant by "no more than one out of ten times will the schedule be shorter/longer than..." If this is done, no additional adjustment in necessary. The procedure in Chapter III can be used to find the absolute limits of the distribution for use in software applications.

If possible, use of a laptop computer or other convenient plotting device should be used to graphically depict the probability distributions that the expert(s) is(are) considering. This will help clear up confusions about the meanings of distribution parameters if done during the interview or group information gathering session.

Investigation of the non-uniform *DPL*© histogram bins illustrated a relationship between the number of histograms (or "intervals" in the *DPL*© set-up menu) and the desired resolution of the attribute under consideration. In general, the maximum range of that attribute from the set of portfolios divided by the number of histogram bins should not be greater than the level of resolution desired. For cost in the preliminary study, the maximum range was a bit over $75M (~ $9M to ~$85M). Since 91 intervals were used

throughout the study, the cost resolution was less than \$1M. Considering the coarseness of the input estimates, this was judged to be sufficient. However, when the decision support system is used with more precise data fed into the LCC Module, the resolution will be much finer. In that case, the number of intervals should be increased appropriately.

The use of a simple point estimate for P(use) is not without hazard. Expression of unknown parameters is preferred to be in terms of probability distributions or intervals, instead of point estimates. Careful sensitivity analysis of this factor is recommended to judge the effects of inaccuracies on the recommended technology portfolios. If the recommendations are very sensitive to a few key estimates of P(use), more effort should be spent on assessing these parameters. Perhaps a panel of experts could be convened to assess these point estimates, using the average of their individual assessments to set the new parameters. If the technology is far enough along in its development cycle, results from developmental tests and evaluations could be used to establish P(use) estimates. Developing historical records concerning P(use) accuracy will be an important consideration.

These techniques are by no means restricted to the DOE technology selection problem. The basic procedure of expressing inputs as random variables and examining the output distributions of relevant decision variables applies to any network of processes.

5.3 *Recommendations for Technology Management and Risk Assessment*

5.3.1 *Sources of Expert Judgement.* Since expert judgement is so critical for technology forecasting, any improvements to the process of soliciting expert opinion will

be of great benefit to the Office of Technology Development. The recently completed tritium study provides an excellent example of what can be done with enough effort. This study compared different tritium production technologies and facility alternatives by pulling together a group of experts and training them in subjective probability estimation to produce cost, schedule, and performance distributions [DOE, 1995e]. Similar training can be given to soil remediation experts brought together at a workshop where, under a group dynamic method such as in Chapter II, release date distributions, probabilities of success in the field, and LCC cost elements can be estimated for a whole group of technologies.

As these emerging technologies move closer to the field, the number of people with sufficient experience with them should grow, making alternative sources of expert opinion easier to find. Other experts besides the technology developers themselves should be cultivated and included in the decision process.

Better surveys and interviews should be designed and refined to solicit assessments from experts. The preliminary questionnaire in Appendix D should be replaced by one that draws on the literature uncovered in this study. Personal interviews, rather than faxed surveys, can improve the acquisition of information by allowing for more interaction and mutual education through interpersonal contact. The additional cost and time required to conduct interviews, however, may dissuade using them for a large group of experts. Interviews allow more data to be collected, including unanticipated information and suggestions, but may result in soliciting estimates from a smaller and potential biased pool

of experts. The trade-offs between desired depth of expert judgement and available resources will have to be made.

DOE policy should require contractors to submit long-term schedules and cost estimates for the development of their products, updating them in annual reporting cycles that are tied to the TTP approval process. Constructing a database of long term schedule and cost estimates at DOE will allow more accountable estimates to be developed. Keeping such a database will help support EM-50's planning and budgeting process. Adherence to these schedules and cost estimates may be a suitable criteria for allocating funding among the development projects.

Using these estimates and documented test results, the accuracy of an expert's predictions over a period of years can be evaluated. From comparisons between actual dates and interim milestone estimates, correction factors for schedule estimates may be empirically developed once sufficient data have been recorded. Requiring the delivery of such historical data is highly encouraged for future technology development contracts written by the Office of Technology Development. Methods beside simple averages can be used to combine different experts' estimates using past accuracies to determine the weights. Selection of the best experts based on past performance will be possible after sufficient records are kept.

Finally, cooperative work with the EPA's SITE program to establish better estimates of probabilities of successful field use can aid EM-50 and EPA as they share test results and collaborate on experiments designed to address the needs of the decision

support system. The impact of incorrect preliminary site characterization can also be investigated.

5.3.2 *Portfolio Management.* Modern investment theory revolves around the concept of managing a group of investments based on the investor's attitudes toward risk and the desired rate of return. The group is viewed as opportunities being created through the investing of resources. A mixture of lower and higher risk investments is sought with the anticipation that some investments will fail. However, these failures are only part of the overall investment, and so no one failure should be devastating. The higher risk investments can provide a better-than-expected return as well as a higher potential for loss [Ryan, 1990:68]. The key is to invest in opportunities whose net incomes are not positively correlated (i.e. all do not lose money at the same time) [Levy and Sarnat, 1990:269].

This idea can be employed by the DOE for managing EM-50's technology development projects. Instead of financial investments, the portfolio consists of technologies, and the opportunities being created are the new capabilities needed for the national remediation effort. A combination of technologies of different levels of expected performance and risk that robustly cover the spectrum of waste types may be a valuable way to manage the risks in the long-term technology development effort.

5.3.3 *Cautions About Risk and Cost-Effectiveness.* New and untried technology is often going to be more inherently risky than older, proven technology. Therefore any technology investment decision based solely on choosing the least risky alternatives is

weighted against selecting emerging technologies. A similar situation is created when comparing life-cycle costs of undeveloped technology, which includes future R&D costs, and established technology, which does not. Inclusion of availability deadlines also creates a situation favoring the old over the new.

While the risk, cost, and availability concerns are valid ones, they must not be the only criteria used. The reason for investing in new technology is to buy future capabilities that are not currently available. This increased expected performance should be included in the decision criteria for technology investment, since it is the primary advantage of emerging technologies. If only the negative aspects of new technologies are measured, the fundamental reason for investing in emerging technology will be neglected.

5.4 *Suggestions for Future Work*

The work in this study can be extended in many directions. One obvious area for further research is the assessment of developmental costs in the decision support system. The current naive uniform annual R&D cost could be replaced by some technology or process-specific cost distribution over the duration of R&D. This would require examining historical cost records and forecasting this shape into the near future. Care would be required, however, to identify and isolate the effects of varying budgetary allocations over the time frame under study.

The model of remediation used in this study relies on the assumption of the independence of individual process durations in the field, given a certain amount of waste to characterize, stabilize, etc. The effects of relaxing this independence assumption would

be a very useful area of study. The individual operational costs and timing of employing a technology in the field could be examined, so to quantify that technology's contribution to the overall portfolio risks.

The expected unfavorable deviations (EUDs) for cost and schedule developed here can be used as independent decision attributes in addition to cost and time as used currently in the DA model. Utility functions for cost and time EUDs could be assessed with DOE technology managers, adding cost and schedule risk explicitly as important decision variables. Mean cost and time for technologies, together with the associated EUDs, could also be used to define a math programming portfolio selection problem, where different combinations of technologies would result from different desired mixtures of risks and expected performance payoffs subject to cost and time constraints [Sherali, et. al., 1994; Weber, et. al., 1990].

Further analysis of the probability of successful implementation of these innovative technologies in the field is warranted. Characterizing this subjective probability through conditional statements of the technology's performance given the presence of specific waste types and items would establish the site-dependent nature of the performance of these technologies. Information from preliminary site assessments could then be used to establish site-specific estimates of the probability of successful use.

While this decision support system is using operations research tools of simulation and decision analysis, this technology selection problem can benefit from other techniques including optimization. Sherali, Alameddine, and Glickman's paper on selecting mixes of

prevention and mitigation alternatives subject to budgetary constraints suggests a way to find an optimally least risky set of new technologies using math programming methods through the concept of risk as undesired events and their likelihoods [1994:197-201]. This treatment of risk, combined with other math programming methods, may allow a different solution technique than the use of *DPL*© simulations.

Concerns about the reaction of stakeholders and public opinion to different remediation technologies was not included in the decision support system. DOE managers do need to take such factors into account in managing emerging technology. Stakeholder values for characteristics of different remediation techniques, such as the use of incineration, on-site disposal, noise and odors given off, could be captured through interviews with cooperative environmental activist organizations and concerned citizen groups. Technologies could then be assigned a general public approval rating that could used in addition to cost, schedule, and performance criteria for decision making.

5.5 *Final Conclusion*

Life-cycle cost analysis and the systematic, quantitative assessment of technical risk are crucial to making good technology management decisions. The techniques described in this study depict technical risk in a simple way, through undesired cost and schedule deviations from expected means, that clearly communicate the basic risks of each alternative remediation strategy to decision makers. It should be remembered that "managers do not enjoy using difficult decision-making methods to make difficult decisions" [Millett and Honton, 1991:74]. In that spirit, explanations of technical risk

should stay simple and concise.

The risks involved in new remediation technology are not the only risks. Programmatic risks have a much greater impact on the overall success or failure of the technology development program than one project's uncertain development schedule. EM-30 and EM-40 remediation efforts that did not use any innovative technology at all still averaged 42% and 18% schedule slippage, respectively, and averaged cost overruns of 48% [DOE, 1993:90, 94, 100].

An effective management cycle of planning, supervising the work, evaluating project status, and reacting with updated plans should be part of technology management practice in EM. If these fundamentals are not present, technology risks are irrelevant since the program will fail in any case. The technology then becomes the scapegoat for the failure of the program [Ryan, 1990:69].

The Department of Energy has no real choice but to manage risk carefully and intelligently. Costs must be controlled and technical risk must be minimized. The methods in this study will provide the DOE with some risk assessment tools required to effectively complete the cleaning up of federal reservations throughout the country.

Appendix A: Notional Technology Data

Technology	Code	Projected Total R&D Costs ($M)	R&D Release Date (years from now)			O&M Costs ($k)			Time to Complete Process (years from now)			P(use)
			earliest	most likely	latest	lowest	most likely	highest	earliest	most likely	latest	
Characterization and Assessment												
Rapid Geophysical Surveyor	ch1	0	0			8.325	20.81	41.63	0.018	0.037	0.074	1
VETEM	ch2	0.235	1	2	4	16.65	41.63	83.25	0.018	0.037	0.074	0.9
High Resolution Imaging	ch3	2	2	3	5	16.65	41.63	83.25	0.018	0.037	0.074	0.9
Stabilization												
In Situ Cementation	s1	0	0			27750	30056	48500	0.163	0.22	0.326	0.95
Innovative Grouting & Retrieval	s2	0.45	4	6	8	14100	15000	22200	0.75	1	2	0.95
In Situ Vitrification	s3	5	1.5	2	4	7500	20000	50000	1.085	1.447	4.34	0.5
Containment												
Monolithic Containment	c1	0	0			2409	4818	9636	0.5	0.75	2	0.8
In Situ Encapsulation	c2	4.03	4	6	8	1445	1499	1927	0.75	1	2	0.9
Soil Saw (Horizontal)	c3	5	3	5	6	12045	14454	19272	0.5	0.75	1	1

Retrieval												
Retrieval Demonstration	r1	0	0			1175	1250	1850	0.161	0.208	0.296	1
Remote Excavation System	r2	0.13	0	0.5	1	1500	2250	4500	0.245	0.316	0.448	0.95
Cooperative Telerobotic Retrieval	r3	10	2.5	3	5	25000	50000	1e+05	0.723	1.447	2.894	0.9
Treatment												
Cementation	t1	0	0			1175	1250	1850	0.081	0.104	0.148	0.85
Plasma Furnace	t2	11	1	2	3	50000	1e+05	3e+05	3.157	6.944	13.89	0.9
STRATEX	t3	0.35	1.5	3	6	3481	3704	5492	1.076	1.389	1.972	0.95
Disposal												
Yucca Mt. Storage Facility	d1	0	12	15	50	71400	84000	88200	0			0.99
On-Site	d2	0	0			3780	3990	4200	0			0.95
Monitoring												
Yucca Mt. Storage Facility	m1	0	12	15	50	0	0	0	0			1
On-Site	m2	0	0			458.4	487.6	721.7	0			1

Table A.1

Data taken from Ralston [1996:110].

Appendix B: Adjusted R&D Release Dates

Process		Dates as Given					Adjusted Dates				R&D cost/yr as given ($k)	R&D adjusted cost/yr ($k)
technology	code	earliest	median	latest	mean	variance	earliest	latest	mean	variance		
Characterization												
Rapid Geophysical Surveyor	ch1	0	0	0	0	0	0	0	0	0	0	0
VETEM	ch2	1	2	4	2.333	0.389	0.549	5.330	2.626	1.001	100.71	89.48
High Resolution Imaging	ch3	2	3	5	3.333	0.389	1.549	6.330	3.626	1.001	600	551.52
Stabilization												
In Situ Cementation	s1	0	0	0	0	0	0	0	0	0	0	0
Innovative Grouting and Retrieval	s2	4	6	8	6	0.667	3.292	9.461	6.251	1.594	75	71.99
In Situ Vitrification	s3	1.5	2	4	2.5	0.292	1.192	5.101	2.764	0.710	2000	1808.75
Containment												
Monolithic Confinement	c1	0	0	0	0	0	0	0	0	0	0	0
In Situ Encapsulation	c2	4	6	8	6	0.667	3.292	9.461	6.251	1.594	716.67	687.89
Soil Saw (Horizontal)	c3	3	5	6	4.667	0.389	2.406	6.937	4.781	0.861	1071.43	1045.81

Retrieval												
Retrieval Demonstration	r1	0	0	0	0	0	0	0	0	0	0	0
Remote Excavation System[3]	r2	0	0.5	1	0.5	0.042	0	1.333	0.611	0.076	260	212.77
Cooperative Telerobotic Retrieval	r3	2.5	3	5	3.5	0.292	2.192	6.101	3.764	0.710	2857.14	2656.51
Treatment												
Cementation	t1	0	0	0	0	0	0	0	0	0	0	0
Plasma Furnace	t2	2	3	4	3	0.167	1.646	4.731	3.126	0.399	3666.67	3519.25
STRATEX	t3	1.5	3	6	3.5	0.875	0.823	7.844	3.889	2.153	100	90
Disposal												
Yucca Mt. off-site[2]	d1	12	15	50	25.67	74.39	8.648	67.61	30.42	174.6	0	0
on-site	d2	0	0	0	0	0	0	0	0	0	0	0
Monitoring												
Yucca Mt. off-site[2]	m1	12	15	50	25.67	74.39	8.648	67.61	30.42	174.6	0	0
on-site	m2	0	0	0	0	0	0	0	0	0	0	0

Table B.1

Notes

1. Technologies that are currently available and have no upcoming release dates have “0” entries.

2. The Yucca Mt. disposal and monitoring option refers to an off-site storage location being considered for the future disposition of radioactive waste. The costs for building this facility will not be paid for out of DOE/EM remediation funds, and so there are no development costs.

3. Since the earliest given date for r2, the Remote Excavation System, is already 0, the standard approach in Chapter 3 cannot be used to find the adjusted latest date. See Appendix E for the additional equations.

Non-Stabilized Portfolios, Basecase

	Cost				
	#1	#2	#3	#4	#5
Mean ($M)	6.56	16.98	18.94	17.01	10.07
Lowest ($M)	3.91	9.16	10.04	14.79	6.19
Highest ($M)	11.77	70.05	85.38	19.7	68.33
Variance ($\$M^2$)	3.91	197.63	205.97	1.47	82.69
Standard Dev. ($M)	1.98	14.05	14.35	1.21	9.09
EUD ($M)	0.7622	5.3341	5.5769	0.4032	2.5535
Semi-variance ($\$M^2$)	2.3646	162.23	164.95	0.7295	75.311
Coef. of Variation	0.3013	0.8277	0.7577	0.0712	0.9027
Norm. EUD	0.1161	0.3141	0.2945	0.0237	0.2535
	Time				
	#1	#2	#3	#4	#5
Mean (years)	3.73	3.14	3.29	5.42	5.29
Lowest (years)	2.3	1.88	1.97	4.08	3.57
Highest (years)	7.65	7.21	7.82	10.47	10.95
Variance ($years^2$)	0.82	1.42	1.43	0.91	1.19
Standard Dev. (years)	0.91	1.19	1.2	0.96	1.09
EUD (years)	0.3452	0.4304	0.4417	0.3717	0.3558
Semi-variance ($years^2$)	0.4835	1.0087	1.0103	0.5686	0.7838
Coef. of Variation	0.2426	0.3797	0.364	0.1764	0.2062
Norm. EUD	0.0925	0.1373	0.1343	0.0686	0.0673

	Total Utility				
	#1	#2	#3	#4	#5
Mean (utility)	0.99379	0.98926	0.98615	0.96184	0.95822
Lowest (utility)	0.77783	0.69768	0.48168	0	0
Highest (utility)	0.99925	0.99932	0.99799	0.995	0.99728
Variance ($utility^2$)	0.00014	0.00055	0.00105	0.00353	0.00674
Standard Dev. (utility)	0.01193	0.02338	0.03247	0.05941	0.08209
EUD (utility)	0.00286	0.00657	0.00826	0.01705	0.02258
Semi-variance ($utility^2$)	0.00013	0.0006	0.00096	0.00309	0.0061
Coef. of Variation	0.012	0.0236	0.0329	0.0618	0.0857
Norm. EUD	0.00288	0.00665	0.00837	0.01773	0.0236

Table C.1

Stabilized Portfolios, Basecase

	Cost				
	#1	#2	#3	#4	#5
Mean ($M)	43.37	39.11	39.08	49.6	49.81
Lowest ($M)	32.7	27.73	29.06	38.23	39.68
Highest ($M)	78.45	71.51	69.93	80.86	79.57
Variance ($\$M^2$)	47.87	39.81	35.08	37.22	33.28
Standard Dev. ($M)	6.92	6.31	5.92	6.1	5.77
EUD ($M)	2.3954	2.2318	2.076	2.0297	1.8861
Semi-variance ($\$M^2$)	31.599	26.148	23.062	24.835	22.096
Coef. of Variation	0.1595	0.1613	0.1516	0.123	0.1158
Norm. EUD	0.0552	0.0571	0.0531	0.0409	0.0379

	Time				
	#1	#2	#3	#4	#5
Mean (years)	1.68	4.01	5.02	5.43	5.48
Lowest (years)	0.92	2.5	3.42	4.08	4.08
Highest (years)	5	7.88	9.61	10.47	10.47
Variance (years^2)	0.47	0.86	0.88	0.91	0.91
Standard Dev. (years)	0.69	0.93	0.94	0.95	0.95
EUD (years)	0.2678	0.35	0.3543	0.3722	0.3734
Semi-variance (years^2)	0.3255	0.5047	0.5195	0.5671	0.5664
Coef. of Variation	0.4084	0.2305	0.1868	0.1759	0.1741
Norm. EUD	0.1593	0.0872	0.0705	0.0686	0.0682
	Total Utility				
	#1	#2	#3	#4	#5
Mean (utility)	0.99184	0.9918	0.98589	0.96986	0.96935
Lowest (utility)	0.7824	0.87073	0.73588	0	0
Highest (utility)	0.99824	0.99812	0.9974	0.99299	0.99275
Variance (utility^2)	0.00016	0.00007	0.00024	0.00236	0.00217
Standard Dev. (utility)	0.01244	0.00849	0.01556	0.04858	0.04654
EUD (utility)	0.00277	0.00243	0.00447	0.00951	0.00914
Semi-variance (utility^2)	0.00014	0.00006	0.00021	0.00221	0.00202
Coef. of Variation	0.0125	0.0086	0.0158	0.0501	0.048
Norm. EUD	0.00279	0.00245	0.00454	0.0098	0.00943

Table C.2

Portfolios After Increasing All P(use) By +10%

	Cost			
	#1 non-stab.	#3 non-stab.	#1 stab.	#3 stab.
Mean ($M)	6.18	13.54	41.91	37.81
Lowest ($M)	3.93	9.75	32.78	29.08
Highest ($M)	11.76	52.61	57.25	52.28
Variance ($\$M^2$)	3	67.01	25.23	18.74
Standard Dev. ($M)	1.73	8.19	5.02	4.33
EUD ($M)	0.5958	1.7861	1.8716	1.6168
Semi-variance ($\$M^2$)	1.8135	63.233	13.163	9.796
Coef. of Variation	0.2804	0.6045	0.1198	0.1145
Norm. EUD	0.0964	0.1319	0.0447	0.0428

	Time			
	#1 non-stab.	#3 non-stab.	#1 stab.	#3 stab.
Mean (years)	3.6	2.97	1.5	4.83
Lowest (years)	2.3	1.97	0.92	3.41
Highest (years)	7.68	7.3	4.08	8.73
Variance (years^2)	0.68	0.88	0.29	0.68
Standard Dev. (years)	0.82	0.94	0.54	0.83
EUD (years)	0.3184	0.2777	0.1697	0.3136
Semi-variance (years^2)	0.3963	0.5314	0.2047	0.3939
Coef. of Variation	0.229	0.3159	0.3591	0.1709
Norm. EUD	0.0885	0.0934	0.1133	0.0646

	Total Utility			
	#1 non-stab.	#3 non-stab.	#1 stab.	#3 stab.
Mean (utility)	0.99518	0.99504	0.99447	0.98944
Lowest (utility)	0.77774	0.80817	0.97648	0.78707
Highest (utility)	0.99925	0.9992	0.99834	0.9974
Variance (utility^2)	0.0000848	0.00029	0.000011	0.00011
Standard Dev. (utility)	0.00921	0.01702	0.00331	0.01048
EUD (utility)	0.002025	0.003005	0.001111	0.002744
Semi-variance (utility^2)	0.0000788	0.0002	0.000008	0.000097
Coef. of Variation	0.0093	0.0171	0.0033	0.0106
Norm. EUD	0.002035	0.00302	0.001117	0.002773

Table C.3

Portfolios After Decreasing All P(use) By -10%

	Cost			
	#1 non-stab.	#3 non-stab.	#1 stab.	#3 stab.
Mean ($M)	6.92	25.77	45.89	41.26
Lowest ($M)	3.81	9.71	32.53	29.01
Highest ($M)	11.77	85.93	78.39	71.23
Variance ($M^2)	4.52	324.79	85.57	62.87
Standard Dev. ($M)	2.13	18.02	9.25	7.93
EUD ($M)	0.91523	7.655578	3.449799	2.696839
Semi-variance ($M^2)	2.562213	216.46	58.34679	42.88073
Coef. of Variation	0.3073	0.06995	0.2016	0.1922
Norm. EUD	0.1322	0.2971	0.0752	0.072

	Time			
	#1 non-stab.	#3 non-stab.	#1 stab.	#3 stab.
Mean (years)	3.88	3.65	1.91	5.25
Lowest (years)	2.3	1.98	0.92	3.42
Highest (years)	7.69	8.08	4.97	9.59
Variance (years^2)	0.93	1.99	0.65	1.04
Standard Dev. (years)	0.97	1.41	0.81	1.02
EUD (years)	0.376964	0.589207	0.339199	0.411124
Semi-variance (years^2)	0.543524	1.3009	0.411465	0.596952
Coef. of Variation	0.2494	0.3866	0.422	0.1939
Norm. EUD	0.0973	0.1616	0.1774	0.0783
	Total Utility			
	#1 non-stab.	#3 non-stab.	#1 stab.	#3 stab.
Mean (utility)	0.99235	0.96974	0.98672	0.97998
Lowest (utility)	0.77569	0.44457	0.78511	0.72902
Highest (utility)	0.99923	0.99803	0.99798	0.99696
Variance (utility^2)	0.000207	0.00374	0.000446	0.000522
Standard Dev. (utility)	0.01439	0.06118	0.02111	0.02285
EUD (utility)	0.003579	0.002025	0.006216	0.007018
Semi-variance (utility^2)	0.000188	0.003305	0.000392	0.000438
Coef. of Variation	0.0145	0.0631	0.0214	0.0233
Norm. EUD	0.003607	0.002089	0.006299	0.007161

Table C.4

Technology Risk Questions

For MSE Interviews

Target Interviewees:	technology developers/principle engineers, first set government project managers, second set waste site managers/owners of the landfill, third set

General Approach:

Always let interviewees explain their answers in their own words – ask for more than just a "yes/no" or number answer.

Make questions as user-friendly as possible.

Leave time for interviewees to add information or additional questions as they see fit.

Include a description of what we mean by terms like "development effort," etc.

Send a letter explaining the purpose of the upcoming interview to the interviewee ahead of time. Include sample questions.

Capt Tom Timmerman, AFIT/ENS
November 22, 1995

Questions for Technology Developers

Terminology:

technology, technical approach: The technology involved with the remediation/characterization product in. All of the product-related issues, including cost, R&D schedule, implementation at a site, etc. is referenced by the "technology" involved.

development effort: The R&D process of developing the technology, starting with concept exploration and going all the way through prototyping and testing. It ends when the technology is ready to be used at a waste site.

implementation: Actual use of the technology at a specific site, with the site manager being the customer. Successful implementation means achieving the remediation goals for that technology, given that the technology was successfully developed.

technology path: The entire set of different technical approaches used in a complete remediation process, starting with characterization of the site and leading through the possible application of stabilization, removal, treatment, disposal, containment, and monitoring technologies.

1. General information
    a. interviewee's name:

    b. name of the project:

    c. TTP number:

    d. name of the DoE manager of the project:

2. Current stage of development

At the time of these answers, where would this development effort fall in the DoE s "technology maturation phases" shown here? [show them the chart]

circle one: basic research, applied research, exploratory development, advanced development, engineering development, demonstration

3. Schedule estimates

a. What is your projected development schedule? May we have a copy of your latest overall schedule?

b. When do you think the technology will be ready for implementation? Could you give a range of dates, including an estimate of lower & upper bounds as well as a most likely date? What are they?

4. Testing & prototypes

Please describe the kinds of testing and demonstrations planned in this development effort, including lab and on-site tests.

5. Mix of proven and emerging technology

a. What kinds of new innovative technology are involved with this technical approach?

b. What relies on proven technology in this technical approach?

c. Please characterize the rough proportion of mature technology vs. emerging technology involved.

6. Budget sensitivity

a. Will you explain how sensitive your development effort is to budget fluctuations from your sponsor? If there was a sudden 10, 25, 50% decrease in your funding, how would that affect the ultimate success of the development? For example, would you be able to continue the project? [-10%, -25%, -50%]

b. How much additional time would be added to the schedule? [-10%, -25%, -50%]

c. Is the project acceptable to your sponsor in such a timeframe? [-10%, -25%, -50%]

7. Applicability

a. What types of waste streams will this technology be applicable to?

i. most effective

ii. effective

iii. minimal effectiveness

iv. no effectiveness

b. Which of the following categories would these waste streams fall into? [volatile organic compounds, semivolatile organic compounds, fuels, inorganics (including radioactives), explosives]

c. What sort of things make up the waste that this technology can handle, e.g. barrels, sludge, liquids, buses, n/a, etc.?

8. R&D costs

a. Could you give an estimate of the range of total expected development costs of this technology, based on the current schedule? Please give a lower and upper bound, as well as a most likely figure.

b. What has been spent on the development up to today? What fraction of the total development has been completed to date?

9. Complexity & Reliability

a. What are the sub-systems involved in this technical approach?

b. What are the expected instrumentation & control costs involved?

10. Secondary wastes and public acceptance

a. What are the expected byproducts or secondary wastes produced using this technical approach at a waste site? What volumes of these byproducts are expected, in relation to the input waste volumes?

b. What sorts of odors, dust, particulates, noise, etc. will be given off?

c. What is the potential for the release of radioactives?

d. What is the potential for operator injury?

11. Interactions with other technologies

a. Are there other characterization/remediation/monitoring technologies that would be well suited to work with this approach in an overall "technology path" treatment of a waste site?

b. Are there other technologies that are required to use this approach?

c. Are there technologies that are incompatible with this one?

11. References

    Would you please list some of your past customers as references?

12. Other

    Is there anything else you'd like to add or comment on?

Questions for Government Managers of Technology Development Projects

Terminology:

technology, technical approach: The technology involved with the remediation/characterization product in. All of the product-related issues, including cost, R&D schedule, implementation at a site, etc. is referenced by the "technology" involved.

development effort: The R&D process of developing the technology, starting with concept exploration and going all the way through prototyping and testing. It ends when the technology is ready to be used at a waste site.

implementation: Actual use of the technology at a specific site, with the site manager being the customer. Successful implementation means achieving the remediation goals for that technology, given that the technology was successfully developed.

technology path: The entire set of different technical approaches used in a complete remediation process, starting with characterization of the site and leading through the possible application of stabilization, removal, treatment, disposal, containment, and monitoring technologies.

1. General information
   a. interviewee's name:

   b. name of the project:

   c. TTP number:

   d. name of the contractor developing the technology:

2. Current stage of development

   At the time of these answers, where would this development effort fall in the DoE s "technology maturation phases" shown here? [show chart]

   circle one: basic research, applied research, exploratory development, advanced development, engineering development, demonstration

3. Schedule
   a. What is the projected development schedule? What fraction of the

total work is complete to date? What fraction of the total development funding has been expended so far?

b. When do you think the technology will be ready for implementation? Could you give a range of dates, including an estimate of lower & upper bounds as well as a most likely date? What are they?

4. Mix of emerging and proven technology

a. Roughly what kinds of new innovative technology are involved with this technical approach?

b. Please characterize the rough proportion of mature vs. emerging technology used.

5. Budget sensitivity

a. Will you explain how sensitive the development effort is to budget fluctuations? If there was a sudden 10, 25, 50% decrease in your funding, how would that affect the ultimate success of the development? For example, would you continue the project? [-10%, -25%, -50%]

b. How much additional time would be added to the schedule? [-10%, -25%, -50%]

c. Is the project acceptable to you in such a timeframe? [-10%, -25%, -50%]

d. Is this project higher priority than the majority of the others being managed by your office, lower priority, or about the same?

e. What kind of budget changes do you anticipate?

6. Applicability

a. What types of waste streams will this technology be applicable to?

i. most effective

ii. effective

iii. minimal effectiveness

iv. no effectiveness

b. Which of the following categories would these waste streams fall into? [volatile organic compounds, semivolatile organic compounds, fuels, inorganics (including radioactives), explosives]

c. What sort of things make up the waste that this technology can handle, e.g. barrels, sludge, liquids, buses, n/a, etc.?

7. R&D costs

a. Could you give an estimate of the range of total expected development costs of this technology, based on the current schedule? Please give a lower and upper bound, as well as a most likely figure.

b. What has been spent on the development up to today? What fraction of the total development has been completed to date?

8. Contractor performance

a. How would you characterize the developer's performance up to now? circle one: excellent, very good, good, fair, poor

b. How have they kept to the original schedule and budget? If there have been changes, why?

9. Secondary wastes and public acceptance

What are the expected byproducts and secondary wastes produced when using this technical approach at a waste site?

10. Contractor references

Can you list some of the contractor's past customers that you know of?

11. Other

Is there anything else you'd like to add or comment on?

Questions for waste site managers

1. Expected landfill contents

a. What volumes of waste do you think are present at your site, using the following categories?

i. volatile organic compounds

ii. semivolatile organic compounds

iii. fuels

iv. inorganics (including radioactives)

1). purely radioactive waste

v. explosives

b. What forms does the waste come in (i.e. sludge, fluids, barrels, boxes, bulky equipment, vehicles, etc.)?

c. How confident are you in the estimate of what waste is in your site? What kind of surprises do you think are likely (i.e. larger/smaller volumes, unexpected waste types, unexpected items, etc.)?

2. Previous site characterizations

a. Has a site characterization ever been done? If so, how was it

conducted? What were the results? Can we get copies of any resulting reports?

b. Is there documentation on what was put into the site and when it was done? If so, may we get copies?

3. Similar sites

Are there any sites that are very similar to yours? What are they?

4. Other

Is there anything else you'd like to add or comment on?

Appendix E: MathCad© *Solution to Release Date Adjustment*

Following the instructions in the MathCad© 5.0+ file, one can convert the expert's estimated triangular release date distribution into the adjusted distribution, to be put into the Technology Database. The following pages show a print-out of this file. To find the adjusted end-points, the appropriate inner fractiles should be entered as indicated. Page E-3 calculates a triangular distribution's mean, variance, PDF, and CDF. In the case where the expert's earliest release date estimate is zero (i.e. the present), use the equations on page E-4.

modified Keefer & Bodily solution method, for x(.03) & x(.90) fractiles

Given an expert's earliest, most likely, and latest estimated release dates, one can solve for the actual earliest and latest dates (when assuming that the expert's dates were really the 3% and 90% interior fractiles, respectively) by putting the expert's estimates in the following three MathCad statements.

expert's earliest date	$x03 := 3$
expert's most likely date	$xm := 5$
expert's latest date	$x90 := 6$

Then, turning on the "SmartMath" option under the "Math" menu above, the Find(x0,x1) statement below will solve the two simultaneous equations under the Given statement.

Given

$$(x03 - x0)^2 = .03\cdot(x1 - x0)\cdot(xm - x0)$$

$$(x1 - x90)^2 = .10\cdot(x1 - x0)\cdot(x1 - xm)$$

One must pick out the feasible pair of bounds from the 4 pairs of solutions below.

$$\mathrm{Find}(x0, x1) \rightarrow \begin{pmatrix} 3.3375895156269938086 & 3.4020090264529935869 & 2.5235299600509455174 & 2.4062636059387435714 \\ 5.6227587950674624771 & 6.7731431572664418002 & 5.5792732197018725849 & 6.9367022226002384634 \end{pmatrix}$$

triangular PDF, mean & variance

formulas taken from Law & Kelton, 1982

earliest date $a := .549$

most likely date $c := 2$

latest date $b := 5.33$

$mean := \frac{a + b + c}{3}$ $mean = 2.626$ $variance := \frac{a^2 + b^2 + c^2 - a \cdot b - a \cdot c - b \cdot c}{18}$ $variance = 1.001$

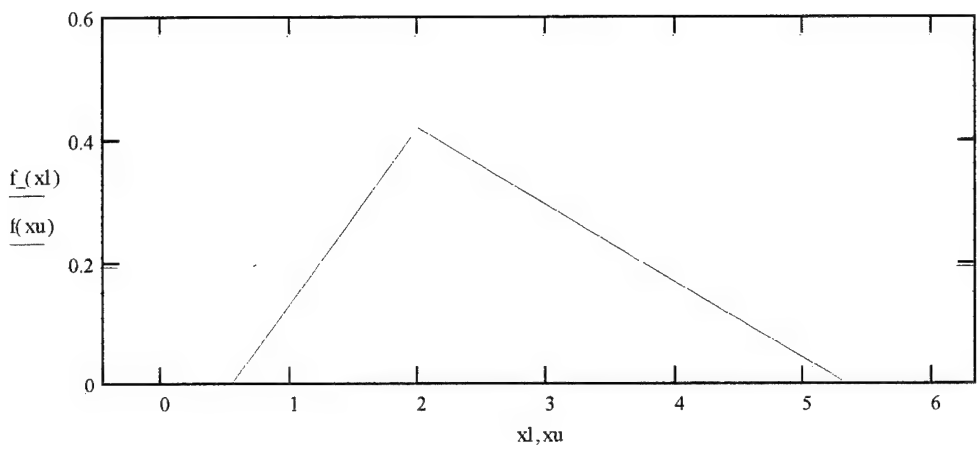
PROBABILITY DISTRIBUTION FUNCTION

$f\_(x) := \frac{2 \cdot (x - a)}{(b - a) \cdot (c - a)}$ (first half of PDF)

$f(x) := \frac{2 \cdot (b - x)}{(b - a) \cdot (b - c)}$ (second half of PDF)

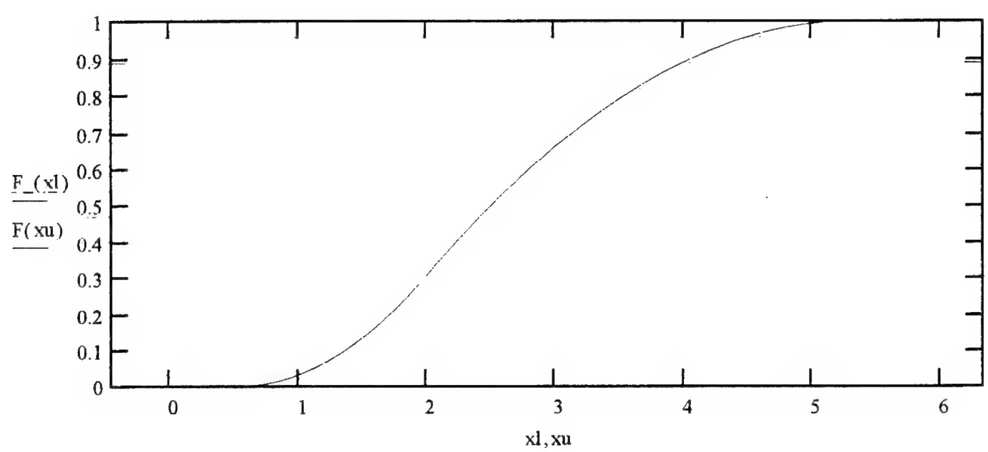
$xl := a, a + .1 .. c$ $xu := c, c + .1 .. b$

These are just counters for the graphs.



CUMULATIVE DISTRIBUTION FUNCTION

(first half of CDF) $F\_(x) := \frac{(x - a)^2}{(b - a) \cdot (c - a)}$ (second half of CDF) $F(x) := 1 - \frac{(b - x)^2}{(b - a) \cdot (b - c)}$



When the expert's earliest release date estimate is 0, the Keefer & Bodily approach breaks down. Use the following equations in that case.

earliest date	$y0 := 0$
expert's most likely date	$ym := .5$
expert's latest date	$y90 := 1$

Then, turning on the "SmartMath" option under the "Math" menu above, the Find(y1) statement below will solve the two simultaneous equations under the Given statement.

Given

$$1 = \frac{1}{2}\cdot\left(\frac{2}{y1}\right)\cdot(y1 - ym) + \frac{1}{2}\cdot\left(\frac{2}{y1}\right)\cdot(ym - y0)$$

$$.1 = \frac{1}{2}\cdot(y1 - y90)\cdot\left[\frac{2}{y1\cdot(ym - y1)} - \frac{2}{ym - y1}\right]$$

One must pick out the feasible upper bound from the pairs of solutions below.

$$\mathrm{Find}(y1) \rightarrow (.83333333333333333333 \quad 1.3333333333333333333)$$

Appendix F: Utility Functions Used in the Pilot Study

General Form

$$u(x) = a + b e^{cx} \tag{F.1}$$

Portfolios Without Stabilization

Cost

	$0 \le$ cost $\le$ \$66M	\$66M < cost
a	1	$-9.58 \cdot 10^{-9}$
b	-0.0001234	121
c	$1.154 \cdot 10^{-7}$	$-7.702 \cdot 10^{-8}$

Time

	$0 \le$ time $\le$ 6.6 yrs	6.6 yrs < time
a	1	$1.066 \cdot 10^{-17}$
b	-0.0001238	121
c	1.153	-0.7702

Portfolios With Stabilization

Cost

	$0 \le$ cost $\le$ \$77M	\$77M < cost
a	1.001	$-2.347 \cdot 10^{-6}$
b	-0.0001273	121
c	$9.852 \cdot 10^{-8}$	$-6.601 \cdot 10^{-8}$

Time

	$0 \le$ time $\le$ 7.7 yrs	7.7 yrs < time
a	1	$2.095 \cdot 10^{-12}$
b	-0.0001245	121
c	0.9879	-0.6601

Appendix G: Non-Uniform DPL© *Histograms*

It is standard practice to use histogram bars of equal width or equal probability, reflecting equal intervals of the attribute in question to collect frequency information. The height of the bar reflects the proportion of the total number of samples that fall inside the interval [Law & Kelton, 1982:180; Mendenhall, et. al., 1990:4].

Many of the histograms resulting from the DA model used in this study have histogram bins of unequal width. Customer service at ADA Decision Systems, the makers of *DPL*©, had no explanation for this behavior. As far as they understood, *DPL*© should produce normal histograms [Dalton, 1996]. The source of this irregularity has not been found at the present time (March 1996).

We have to consider the possibility that the irregularity is caused by some error in *DPL*©. The effect of this irregular bin sizing would then introduce further error into calculations of the mean, variance, and EUD with Equations 3.4, 3.6, and 3.7. In this case, instead of representing bin members by the midpoints of equally sized bins, the midpoints of larger width bins give less weight to their members than those of narrow bins. Since potentially three or four narrow bars might fit inside a wide bar, the wider bin midpoint counts a third or fourth as much as the ones from the narrower bins.

This additional error emphasizes the fact that these histograms and all the statistics drawn from them are approximations of sample characteristics, which are themselves estimates of population characteristics. Fortunately, as the number of iterations for each

run of the DA model used here (10,000) is high enough to support the use of the central limit theorem in establishing approximate confidence intervals and testing hypotheses about the sample means [Mendenhall, et. al., 1990:319].

To indirectly examine the effect of the non-uniform histogram bins, the number of intervals *DPL*© uses to collect the histogram data was increased from the default value of 91 to 1488, the maximum available. While there are still histogram bins of unequal size in the 1488 case, there are much fewer and they carry less weight. The non-stabilized #3 portfolio was used. The means, variance, and EUDs of the two runs are summarized in Table G.1.

Comparison of Cost Results for 1488 vs. 91 Histogram Intervals

for the #3 portfolio w/o stabilization

	Mean ($M)	Variance $(\$M)^2$	EUD ($M)
1 - 91 intervals	18.94	205.97	5.577
2 - 1488 intervals	19.029	206.77	5.624

Table G.1

Using the same procedures described in section 4.2.6.2 in Chapter 4, we can test the hypotheses that the population means and variances that underlie these results are the same.

The test for the equality of the variances uses a test statistic of $\hat{F} = \frac{S_2^2}{S_1^2}$ (since $S_2^2 > S_1^2$). Again, because F statistics tables and software do not include degrees of freedom as high as 10,000/10,000, we need to look at a bound of $F_{\frac{\alpha}{2}, 1000, 1000}$. At an α of 0.01,

the rejection region threshold is 1.18. Since $\frac{206.77}{205.97} = 1.003884$, we fail to reject the hypothesis that the two means are equal (the necessary p-level to reject the null hypothesis is 0.23779).

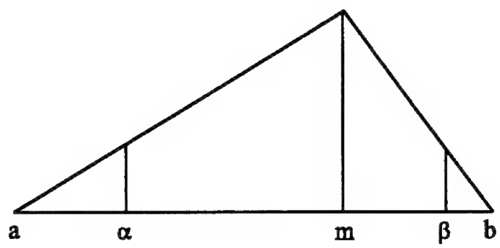
The test for the equality of the means uses a test statistic of $\hat{t} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ and a rejection region of 2.765 for an $\alpha = 0.01$. In this case our test statistic is $4.381 \cdot 10^{-7}$, which certainly does not fall inside the rejection region of greater than 2.765. At the 99% significance level, we fail to reject the null hypothesis of the populations means being equal, assuming the two distributions are normal. Even though the assumption is not a good one, this result supports the continued use of the irregular *DPL*© histograms.

Appendix H: Hudak's Adjustment to Triangular Distributions

Hudak, in his 1994 article "Adjusting Triangular Distributions for Judgemental Bias," describes a way to find the endpoints of a triangular distribution given the mode and two interior fractiles. This appendix provides the core of his method [1994:1027].

The right end point, b, can be found with the solution to the following four-degree polynomial:

$$d_1b^4 + d_2b^3 + d_3b^2 + d_4b + d_5 = 0$$



where

X = x^{th} fractile as a fraction (i.e. $X = 0.1$ for the 10^{th} percentile)

$Y = y^{th}$ fractile as a fraction (i.e. $Y = 0.9$ for the 90^{th} percentile)

$Z = 1 - Y$

$\alpha = x^{th}$ fractile [given]

$\beta = y^{th}$ fractile [given]

m = mode [given],

and

$d_1 = a_1^2 - c_1$

$d_2 = 2a_1a_2 - c_2$

$d_3 = 2a_1a_3 + a_2^2 - c_3$

$d_4 = 2a_2a_3 - c_4$

$d_5 = a_3^2 - c_5$

$a_1 = 1 - Z$

$a_2 = Z\alpha + Zm - 2\beta$

$a_3 = \beta^2 - Z\alpha m$

$c_1 = X(1 - Z)$

$c_2 = X(2Zm - (4 - 2Z)\beta)$

$c_3 = X((6 - Z)\beta^2 - 4Z\beta m - Zm^2)$

$c_4 = X(-4\beta^3 + 2Z\beta^2 m + 2Z\beta m^2)$

$c_5 = X(\beta^4 - Z\beta^2 m^2)$

Once b has been determined, find a with:

$$a = b - (b - \beta)^2 / (Z(b - m))$$

The solution to the four-degree polynomial will involve four real roots. The resulting pairs of b and a must be checked against β and α — only one pair will satisfy the restrictions on a and b ($a < \alpha$, $b > \beta$).

That pair are the endpoints to the triangular distribution, and will fully specify it together with m.

Bibliography

Ashford, R. W., R. G. Dyson, and S. G. Hodges. "The Capital-Investment Approach of New Technology: Problems, Misconceptions, and Research Directions," *Journal of the Operational Research Society:* v. 39, n. 7 (July 1988). p. 637-42.

Ashton, Robert H. "Combining the Judgements of Experts: How Many and Which Ones?," *Organizational Behavior and Human Decision Processes:* v. 38. p. 405-14.

Attaway, L. D. "Criteria and the Measurement of Effectiveness," *Systems Analysis and Policy Planning: Applications in Defense*. E. S. Quade and W. I. Boucher, ed. RAND report R-439-PR (abridged). New York: Elsevier, 1968.

Auclair, Lt Col Paul. AFIT/ENS department head. Personal discussions. AFIT, Wright-Patterson AFB, OH. 19 Jan 96.

Biery, Frederick P. "The Accuracy of Military Cost and Schedule Forecasts," *Journal of Cost Analysis:* 3, Spring 1986. p. 13-23.

-----, David Hudak, and Shishu Gupta. "Improving Cost Risk Analyses," *Journal of Cost Analysis:* Spring, 1994. p. 57-85.

Bhat, Vasanthakumar. "Generic Risk Reduction Strategies for R&D Projects," *Technology Management: Proceedings of Portland International Conference on the Management of Engineering and Technology*. Kocaoglu & Niwa, ed. Portland OR: Portland State University, 1991. p. 262-5.

Bronson, Po. "George Gilder," *Wired:* v. 4, n. 3 (March 1996). p. 122-7, 186-95.

Bunn, D. W. "Discussion of Paper by Dr Newbold and Prof. Granger," *J. R. Statist. Soc. A:* 137, part 2 (1974). p. 158-9.

Cain, Joseph P. Class lecture, OPER 632. Air Force Institute of Technology, Wright-Patterson AFB OH. 5 Feb 1996.

Chicken, John C. *Managing Risks and Decisions in Major Projects*. London: Chapman & Hall, 1994.

Choobiney, F., & A. Behrens. "Use of Intervals & Possibility Distributions in Economic Analysis," *Journal of Operations Research Society:* v. 43, n. 9 (Sept 92). p. 907-18.

Coleman, Cynthia-Lou. "Science, Technology, and Risk Coverage of a Community Conflict," *Media, Culture, and Society:* v. 17 (1995). p. 65-79.

Clemen, Robert T. *Making Hard Decisions: An Introduction to Decision Analysis*. Belmont, CA: Duxbury Press, 1991.

Dalton, Chris. ADA technical support. Personal communication. 12 Feb 1996.

Deckro, D. F., J. A. Jackson, & J. M. Kloeber. *Life Cycle Costing & Decision Analysis for Landfill Technology Assessment.* AFIT/ENS proposal. Wright-Patterson AFB OH: Operational Sciences Department, Air Force Institute of Technology, 18 Mar 95.

Defense Systems Management College. *Risk Management: Concepts and Guidance*. Washington, D.C.: GPO, 1989.

Department of Defense, Environmental Technology Transfer Committee. *Remediation Technologies Screening Matrix and Reference Guide*, 2 ed. Aberdeen Proving Ground, MD: US Army Environmental Center, October 1994.

Department of Energy, Offices of Environmental Restoration and Waste Management. *Project Performance Study*. Reston, VA: Independent Project Analysis, Inc., 30 November 1993.

Department of Energy, Office of Technology Development. *FY-95 Technology Catalog: Technology Development for Buried Waste Remediation*. DOE/ID-10513. Washington D.C.: Office of Technology Development, 1995.

-----. *Landfill Stabilization Focus Area Technology Summary*. DOE/EM-0251. Washington D.C.: Office of Technology Development, June 1995.

Department of Energy, Office of Environmental Management. *Risks and the Risk Debate: Searching for Common Ground, "The First Step."* DRAFT report. Vol. I. June 1995.

Department of Energy, Office of Reconfiguration. *Final Programmatic Environmental Impact Statement for Tritium Supply and Recycling*, vol. II. DOE/ESI-0161. October 1995.

Devore, Jay L. *Probability and Statistics for Engineering and the Sciences*, 2nd ed. Monterey, CA: Brooks/Cole Publishing Co., 1987.

-----. *Technical Reference Report for Tritium Supply and Recycling*. DOE/DP-0134. Washington D.C.: Office of Reconfiguration, DP-25, October 1995.

Feizollahi, Fred, and William J. Quipp. *Integrated Thermal Treatment System Study: Phase 2 Results*. INEL-95/0129. Idaho Falls, ID: Lockheed Idaho Technologies Company, August 1995.

Gupta, Sunil, and Peter C. Wilton. "Combination of Forecasts: An Extension," *Management Science:* v. 33, n. 3 (March 1987). p. 356-72.

Haimes, Yacov Y., Timothy Barry, and James H. Lambert, ed. "When and How Can You Specify a Probability Distribution When You Don't Know How," *Risk Analysis:* v. 14, n. 5 (October 1994). p. 661-706.

Harmon, Marion. *First There Was ISO 9000, Now There is . . . ISO 14000.* Quality Digest on-line information service ((800)527-8875). 1994.

Helton, Jon C. "Treatment of Uncertainty in Performance Assessments of Complex Systems," *Risk Analysis:* v. 14, n. 4 (Aug 94). p. 483-511.

Hudak, David G. "Adjusting Triangular Distributions for Judgemental Bias," *Risk Analysis:* v. 14, n. 6 (Dec 94). p. 1025-31.

Jackson, Jack A., Thomas P. White, Jack M. Kloeber, Ronald J. Toland, Joseph P. Cain, and Dorian Y. Buitrago. *Comparative Life-Cycle Cost Analysis for Low-Level, Mixed Waste Remediation Alternatives*. AFIT Technical Report 95-01. Wright-Patterson AFB, OH: Air Force Institute of Technology, March 1995.

Jia, Jianmin, and James S. Dyer. *A Standard Measure of Risk and Risk-Value Models*. Risk-Value Study Series, Working Paper no. 1. Austin, TX: Dept. of Management Science and Information Systems, University of Texas -Austin, October 1995.

Keefer, Donald L., and Samuel E. Bodily. "Three-Point Approximations for Continuous Random Variables," *Management Science:* v. 29, n. 5 (May 1983). p. 595-609.

Keeney, Ralph L., & Howard Raiffa. *Decisions With Multiple Objectives: Preferences and Value Trade-offs*. New York: J. Wiley & Sons, 1976.

King, William R., and Talmadge A. Wilson. "Subjective Time Estimates in Critical Path Planning: A Preliminary Analysis," *Management Science:* v. 13, n. 5 (January 1967). p. 307-20.

-----, Donald M. Wittevrongel, and Karl D. Hezel. "On the Analysis of Critical Path Time Estimating Behavior," *Management Science:* v. 14, n. 1 (September 1967). p. 79-84.

Klein, Jonathan H. "Modeling Risk Trade-Off," *Journal of the Operational Research Society:* v. 44, n. 5 (May 1993). p. 445-60.

Law, Averill M., and W. David Kelton. *Simulation, Modeling, and Analysis*. New York: McGraw-Hill Book Co., 1982.

-----. *Simulation, Modeling, and Analysis*, 2nd ed.. New York: McGraw-Hill Book Co., 1991.

Leonard-Barton, Dorothy. *Implementing New Technology: The Transfer from Developers to Operations*. Working paper 87-049. Division of Research, Harvard Business School, June 1987.

Levary, Reuven R., and Neil E. Seitz. *Quantitative Methods for Capital Budgeting*. Cincinnati, OH: South-Western Publishing Co., 1990.

Levy, Haim, and Marshall Sarnat. *Capital Investment and Financial Decisions*, 4th ed. New York: Prentice Hall, 1990.

Luftig, Steven. EPA. Address to the International Containment Technology Workshop. Omni Hotel, Baltimore MD. 29 Aug 95.

Makridakis, Spyros, and Robert L. Winkler. "Averages of Forecasts: Some Empirical Results," *Management Science:* v. 29, n. 9 (1983). p. 987-96.

-----, Steven C. Wheelwright, and Victor E. McGee. *Forecasting: Methods and Applications*. New York: J. Wiley & Sons, 1983.

Mendenhall, William, Dennis D. Wackerly, and Richard L. Scheaffer. *Mathematical Statistics With Applications*. Belmont, CA: Duxbury Press, 1990.

Meredith, Jack R., & Samuel J. Mantel. *Project Management: A Managerial Approach*. NY: J. Wiley, 1995.

Mitchell, Graham R. "Alternative Frameworks for Technology Strategy," *European Journal of Operations Research:* v. 47, no. 2 (25 Jul 90). pp. 153-161.

Millett, Stephan M. & Edward J. Honton. *A Manager's Guide to Technology Forecasting and Strategic Analysis Methods*. Columbus, OH: Battelle Press, 1991.

Mohuidden, Dr Jaffir. Landfill Stabilization Focus Area Manager, EPA/EM-55. Project discussions with AFIT personnel. Waste Policy Institute offices, Gaithersburg, MD. 18 Aug 1995.

-----. Address to the International Containment Technology Workshop. Omni Hotel, Baltimore MD. 29 Aug 1995.

Morris, Peter A. "Combining Expert Judgements: A Bayesian Approach," *Management Science:* v. 23, n. 7 (March 1977). p. 679-93.

Mykytka, Edward. AFIT/ENS professor. Personal electronic mail communication. 13 Feb 1996.

-----. Personal electronic mail communication. 20 Feb 1996.

Newbold, P., and C. W. J. Granger. "Experience with Forecasting Univariate Time Series and the Combination of Forecasts," *J. R. Statist. Soc. A:* 137, part 2 (1974). p. 131-46.

Nickelson, Dave. Personal communication (fax) to Steve Antonielli. AFIT/ENS, Wright-Patterson AFB, OH. 23 Jan 96.

Porter, Alan L., A. Thomas Roper, Thomas W. Mason, Frederick A. Rossini, and Bradley J. Wiederholt. *Forecasting and Management of Technology*. New York: J. Wiley & Sons, 1991.

Priestly, M. B. "Discussion of Paper by Dr Newbold and Prof. Granger," *J. R. Statist. Soc. A:* 137, part 2 (1974). p. 152-3.

Ralston, Brian E. *Development of a Decision Support System For the Department of Energy's Selection of Waste Site Remediation Technologies*. MS thesis, AFIT/GOR/ENS/96M-12. Graduate School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1996.

Rumer, Ralph R., & Michael E. Ryan, ed. *Barrier Containment Technologies for Environmental Remediation Applications*. NY: J. Wiley & Sons, 1995.

Ryan, Hugh W. "Assessing the Risk of New Technology," *Journal of Information Systems Management:* Winter, 1990. p. 68-72.

Sackman, H. *Delphi Assessment: Expert Opinion, Forecasting, and Group Process*. RAND report R-1283-PR. Santa Monica, CA: RAND Corporation, April 1974.

Sherali, Hanif D., Amine Alameddine, and Theodore S. Glickman. "Biconvex Models and Algorithms for Risk Management Problems," *American Journal of Mathematical and Management Sciences*: v. 14, n. 3/4 (1994). p. 197-228.

Shina, Sammy G. *Concurrent Engineering and Design for Manufacture of Electronics Products*. New York: Van Nostrand Reinhold, 1991.

Troxler, J. W., and P. L. Schillings. "Another Look at Uncertain Risk Management When Evaluating Manufacturing Technologies," *Computer-Integrated Manufacturing Systems:* v. 6, n. 4 (November 1993). p. 229-34.

VanHorne, James C. *Financial Management and Policy*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1971.

Weber, R., B. Werners, & H.-J. Zimmerman. "Planning Models for Research & Development," *European Journal of OR:* v. 48 (1990). p. 175-88.

Wheeler, Charles M. *Risk Perception and Technology: The Sense of the Irrational.* University of North Carolina at Greensboro: Greensboro, NC, 16 July 1993.

White, Thomas P., Ronald J. Toland, and Dorian Y. Buitrago. *Life-Cycle Cost Analysis for Radioactive Waste Remediation Alternatives*. MS thesis, AFIT/GOA/GOR/ENS/95M-50. Graduate School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1995 (AD-AAL7137).

Williams, T. M. "Practical Use of Distributions in Network Analysis," Journal of the Operational Research Society: v. 43, n. 39 (1992). P. 265-70.

Winkler, Robert L. "Combining Probability Distributions From Dependent Information Sources," *Management Science:* v. 27, n. 4 (April 1981). p. 479-88.